

INSTITUT FÜR INFORMATIK
UND PRAKTISCHE MATHEMATIK

**Multidimensional Wavelets and
Computer Vision**

Kai Neckels

Bericht Nr. 0303

Mai 2003



CHRISTIAN-ALBRECHTS-UNIVERSITÄT
KIEL

Institut für Informatik und Praktische Mathematik der
Christian-Albrechts-Universität zu Kiel
Olshausenstr. 40
D – 24098 Kiel

Multidimensional Wavelets and Computer Vision

Kai Neckels

Bericht Nr. 0303
Mai 2003

e-mail: kn@ks.informatik.uni-kiel.de

Dieser Bericht enthält die Dissertation des Verfassers

1. Gutachter

Prof. G. Sommer (Kiel)

2. Gutachter

Prof. B. Fischer (Lübeck)

3. Gutachter/Beisitzer

Prof. U. Heute (Kiel)

Datum der mündlichen Prüfung

24.04.2003

Für Silvia

Acknowledgments

This work would have never been completed without the help and support of several people. First of all, I want to thank my supervisor Prof. Dr. GERALD SOMMER. He was the one to give me the chance to do my research in the area of computer vision and he also provided me with the first ideas for the composition of this doctoral thesis. I also have to thank him for leaving me the great freedom to choose the concrete topic of my research as I thought it would be the best and also for letting me do this work in the way I prefer to work. His motivations and fruitful hints also had a striking impact on the results in this work.

Furthermore, I am grateful to Prof. Dr. BERND FISCHER from the Medical University of Lübeck for being my second advisor and for giving me the occasion to discuss my research work with him. I also want to thank the rest of the Cognitive Science research group at the University of Kiel, my colleagues as well as the technical staff, who helped whenever I came up with any kind of problem. Special thanks go to our secretary, FRANCOISE MAILLARD, for being always helpful and doing all these little things that are so important to keep the work going on and to my office-mates BODO ROSENHAHN and YI-WEN ZHANG for friendly taking me in *their* office and to answer numberless questions. I am also grateful to MARKUS VON OEHSSEN for some helpful discussions and for providing me some of his MatLab code.

This work was granted by the “Deutsche Forschungsgemeinschaft” (DFG) within the “Graduiertenkolleg Effiziente Algorithmen und Mehrskalenmethoden” (GEAM) at the Christian-Albrechts-Universität Kiel. I want to thank these institutions for the financial support and for maintaining the required infrastructure that is essentially necessary to do this kind of research.

The experiments carried out in this work were done under extensive usage of the software packages Maple[®] and MatLab[®]. I am thankful to the authors for providing such excellent mathematical tools — these helped to improve several of my research results. This thesis was written with the help of the T_EX and L^AT_EX systems by DONALD E. KNUTH and LESLIE LAMPORT, whom I also have to thank very much — they made this work look good.

I want to thank my parents for teaching me the important things of life and for all their guidance, support and love they gave me over the years. Without their help, nothing of this would have been possible at all. Finally, I want to thank my girl friend SILVIJA VUKOVAC, a person who has absolutely nothing to do with wavelets or computer vision but who nevertheless has the most impact on the fact that this work was ever accomplished. She was always there when I needed her support and gave me the strength to continue my work even in rainy days — she gets all my love.

Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Konstruktion und der mathematischen Analyse von mehrdimensionalen, nicht separablen Wavelets und deren effizienter Anwendung im Bereich Computer Vision. Wavelets sind eine erst relativ kurz bekannte Klasse von mathematischen Funktionen mit einer Reihe charakteristischer Eigenschaften, welche sie für den Einsatz in der numerischen Signalverarbeitung besonders attraktiv machen. Zu den wesentlichen Eigenschaften zählen hierbei die Zugehörigkeit zum HILBERTraum $L_2(\Omega)$, ihre Skalierbarkeit, das Besitzen verschwindender Momente, eine Basis- oder Frame-Eigenschaft und im besonderen die Tatsache, dass sie eine gute Lokalisierungseigenschaft sowohl im Zeit- als auch im Frequenzraum besitzen. Aus diesen Merkmalen lassen sich speziell zwei fundamentale Konzepte zur Verwendung von Wavelets ableiten: Zum einen ist dies die Benutzung von Wavelet-Basen zur *Repräsentation* von mehrdimensionalen Daten und zum anderen handelt es sich um die effiziente Beschreibung von *Operatoren* in Wavelet-Basen. Aus der Verbindung dieser beiden Konzepte lassen sich schnelle Algorithmen für eine Vielzahl von Anwendungen erzeugen.

Diese Dissertation besteht im wesentlichen aus drei Teilen. Der erste Teil erörtert die fundamentalen Prinzipien, die dem Entwurf mehrdimensionaler Wavelet-Filter zu Grunde liegen. Hier geht es zum einen um die Wahl der im Mehrdimensionalen nicht mehr trivialen (geometrischen) Form der Unterabtastung des zu bearbeitenden Signals. Diese Fragestellung wird in einer eingehenden Untersuchung mehrdimensionaler Skalierungsmatrizen im Kapitel 3 beantwortet. Das zweite, wesentlich fundamentalere Problem bei der Erzeugung mehrdimensionaler Wavelets ist das Problem des Filter-Designs an sich. Es wird hierbei gezeigt werden, dass (eindimensionale) Wavelet-Filter typischerweise mittels Spektralfaktorisierung trigonometrischer Polynome erzeugt werden können. Da in höheren Dimensionen jedoch keine generellen Faktorisierungsaussagen getroffen werden können, ist die Verwendung alternativer

Konzepte gefragt. Im Verlaufe des ersten Teils dieser Arbeit werden verschiedene Filter-Design-Verfahren für mehrdimensionale Wavelets hergeleitet werden.

Im zweiten Teil geht es um die Untersuchung der analytischen Eigenschaften mehrdimensionaler Wavelets. Es werden verschiedene Approximationsaussagen und die schnelle Wavelet-Transformation für mehrdimensionale Signale bereit gestellt. Anschließend wird die effiziente Repräsentation von Operatoren in mehrdimensionalen Wavelet-Basen untersucht und einige Beispiele mit Bezug zu Computer Vision Anwendungen, wie Differential- oder Translationsoperatoren, werden ausführlich diskutiert. Um mit den Operatoren in dieser Form (numerisch) arbeiten zu können, werden in der Folge verschiedene Methoden zur Wavelet basierten Lösung inverser Probleme behandelt. Hierbei wird sich zeigen, dass sowohl die intrinsischen (durch den Filter determinierten) Eigenschaften der Wavelets als auch deren generelle Approximations- und Frame-Eigenschaften von fundamentaler Bedeutung für die Qualität der entsprechenden Algorithmen sind.

Teil drei der Arbeit ist konkreten Computer Vision Anwendungen gewidmet. An Hand dieser Anwendungen soll die Umsetzung der zuvor entwickelten Theorie in die Praxis demonstriert werden. Zunächst werden lineare und nicht lineare Skalenraumkonzepte vorgestellt, im Anschluß daran geht es um die (Wavelet basierte) Berechnung von optischem Fluss in Bildsequenzen. Skalenraumtheorien sind im Hinblick auf Computer Vision für sich selbst ein interessantes Studienobjekt, da sie ein hierarchisches Bildmodell über einem freien Skalenparameter, der dem Fokus eines natürlichen Beobachters nahe kommt, entwickeln. Insbesondere nicht lineare Skalenräume eignen sich aber auch im Kontext anderer Computer Vision Anwendungen zur Regularisierung von unerwünschten Effekten wie der Glättung von signifikanten Kanten. Diese Eigenschaft wird bei der Berechnung von optischem Fluß zur besseren Herausbildung von Bewegungskanten von hohem Nutzen sein. Im Verlaufe dieses dritten Teils der Arbeit wird sich zeigen, wie die zuvor entwickelten Wavelet basierten Konzepte zur effizienten und zuverlässigen Lösung der genannten Computer Vision Probleme eingesetzt werden können.

Naturgemäß sind in einer wissenschaftlichen Forschungsarbeit nicht ausschließlich neue Ideen und Theorien zu finden, sondern auch viel grundlegendes Material zur Verbesserung des Verständnisses. In dieser Dissertationsschrift wurde versucht, fremde Quellen und Resultate so weit wie möglich anzugeben, was in einem recht umfangreichen Literaturverzeichnis seinen Ausdruck findet. Um die eigenen Beiträge des Autors besser hervorzuheben, wurden die entsprechend gefassten mathematischen Aussagen mit einem hochgestellten Stern versehen (z.B. "LEMMA 5.8*" o.Ä.). Eine derartige Kennzeichnung in

laufendem Text oder Formeln ist selbstverständlich nicht möglich — jedoch sollte aus den entsprechenden Formulierungen stets klar werden, wann es sich um Resultate handelt, die nicht auf den Autor zurückgehen.

Summary

The present work deals with the construction and the mathematical analysis of multidimensional, nonseparable wavelets and their efficient application in the area of computer vision. Wavelets are a rather newly known class of mathematical functions with a set of characteristic properties, which make them especially attractive for the usage in numerical signal processing. Some of the essential properties are the membership in the HILBERT space $L_2(\Omega)$, their scalability, the possessing of vanishing moments, a basis or frame property and especially the fact, that they are well localized in the spatial as well as in the frequency domain. There are two very special fundamental concepts to be derived out of this features: For the first, this is the usage of wavelet bases to *represent* multidimensional data and for the second, this is the efficient description of *operators* by means of waveletbases. The conjunction of these both concepts leads to fast algorithms for a number of concrete applications.

This thesis consists mainly out of three parts. The first part discusses the fundamental principles, that lie at the heart of multidimensional wavelet filter design. Hereby, one has to get along with the choice of the in higher dimensions no longer trivial (geometric) shape of oversampling of the signal to be processed. This question will be answered in a thorough investigation of multidimensional dilation matrices in Chapter 3. The second, much more fundamental problem in the creation of multidimensional wavelets is the filter design problem itself. It will be shown that (onedimensional) wavelet filters are typically produced via spectral factorization of certain trigonometric polynomials. But since there exist no general factorization theorems in higher dimensions, the application of alternative concepts is required. During the running of the first part of this work, various filter design methods for multidimensional wavelets are developed.

In the second part, the analytical properties of multidimensional wavelets are investigated. Various approximation results and the fast wavelet transform are

provided. In the sequel, the efficient representation of operators in bases of multidimensional wavelets is investigated and some examples related to computer vision applications, like differential or translation operators, are discussed in some detail. In order to work with these form of operator representation, varied methods for the wavelet based solution of inverse problems are investigated afterwards. It will turn out that the intrinsic properties (which are determined by the filter) as well as the general approximation and frame properties of wavelets are of fundamental significance for the quality of the corresponding algorithms.

Part three of this work is dedicated to concrete computer vision applications. By means of these applications, the realization of the developed theory in a practical surrounding shall be demonstrated. First, linear and nonlinear scale space concepts are introduced, followed by the discussion of (wavelet based) optical flow estimation in image sequences. In regard of computer vision, scale space theories are an interesting subject for themselves, since they produce a hierarchical image model with a free scaling parameter that corresponds in some sense to the focus of a natural observer. But especially nonlinear scale spaces are also very well-suited in the context of other computer vision applications as regularizers for undesired by-effects like smoothing across significant edges. This property will play a prominent role in the motion boundary recovery within the optical flow estimations and lead to noteworthy improvements. During the running of this third part of the work, it will turn out how the developed wavelet based concepts and theories can be applied in order to obtain efficient and reliable solutions to the abovementioned computer vision problems.

Naturally, there are not only new results and theories included in a scientific research work, but also lots of basic material to assure a better understanding of the complete text. In this thesis, the author tried to quote external sources and results as far as possible; this fact finds its expression in a very comprehensive reference list. To better emphasize the original contributions of the author, the accordingly captured mathematical statements were marked by a high-positioned star (e.g. “LEMMA 5.8*” or similarly). A corresponding marking would be obviously unfeasible during the running text, but it should always be clear from of the formulations, if a certain result is not due to the author.

Contents

Acknowledgments	v
Zusammenfassung	vii
Summary	xi
List of Figures	xvii
List of Tables	xix
Notation	xxi
1. OUVERTURE	1
1.1 A Brief History of Wavelets	2
1.2 About Computer Vision	6
1.3 Organization of the Thesis	8
Part I Multidimensional Wavelets	
2. ONEDIMENSIONAL WAVELETS	13
2.1 Wavelet Paradigms	14
2.2 Construction of Onedimensional Wavelets	18
2.3 Related Work	26
3. MRA AND SAMPLING IN \mathbb{R}^n	29
3.1 Multiresolution Analyses in Higher Dimensions	30
3.2 Uniform Affine Sampling Lattices in \mathbb{R}^2	31
3.2.1 Regular Sampling	32
3.2.2 The Quincunx Lattice	33
3.2.3 The Hexagonal Case	38
3.3 Higher Dimensions	41
3.3.1 Classification of Dilation Matrices	42
3.3.2 The Isotropy Measure ζ_n for Increasing n	45
3.4 Extensions and Related Work	48
4. DESIGNING WAVELETS	51
4.1 Conceptual Generalization	52

4.1.1	Some Fundamental Relations	52
4.1.2	Basic Filters in Higher Dimensions	55
4.1.3	Why Using Nonseparable Multidimensional Wavelets?	56
4.2	Orthogonal Wavelets in \mathbb{R}^n	58
4.2.1	General Design Principles	59
4.2.2	A First Example	62
4.2.3	Coiflets	64
4.2.4	Filter Optimization	65
4.2.5	Other Approaches	72
4.3	Biorthogonal Wavelets in \mathbb{R}^n	74
4.4	Tight Wavelet Frames	76
4.4.1	Introduction	76
4.4.2	A New and Easy Way to Build Tight Wavelet Frames	78
Part II Wavelet Analysis in Higher Dimensions		
5.	BASICS	87
5.1	The Projection Operators \mathbf{P}_j	87
5.2	The Fast Matrix Dilation Wavelet Transform	90
5.3	Generalized Wavelet Approximation	92
6.	OPERATORS IN BASES OF NONSEPARABLE WAVELETS	97
6.1	Differential Operators and Connection Coefficients	98
6.1.1	Definition and Fundamental Properties	98
6.1.2	Calculation of Fundamental Connection Coefficients	101
6.1.3	Relations to Finite Difference Schemes	105
6.2	Affine Transforms in Wavelet Domain	108
6.3	Wavelet Representations of Operators	111
6.3.1	The Standard Form	111
6.3.2	The Non-Standard Form	112
6.4	Related Work	114
7.	ILL-POSED PROBLEMS	117
7.1	Problem Statement	117
7.1.1	The Galerkin Method	118
7.2	Numerical Treatments	119
7.2.1	Multiscale Preconditioning	120
7.2.2	Time Discretization of (Possibly) Nonlinear Operators	123
7.2.3	Additive Operator Splitting	124
7.3	Related Work	127

Part III Multidimensional Wavelets in Computer Vision

8. LINEAR AND NONLINEAR SCALE SPACES	131
8.1 Linear Scale Spaces	132
8.1.1 The Heat Conduction Equation	133
8.1.2 Linear Wavelet Scale Space	134
8.2 Nonlinear Scale Space Extensions	136
8.2.1 The Perona-Malik Approach	136
8.2.2 Anisotropic Diffusion	140
8.2.3 Nonlinear Wavelet Estimation	142
8.3 Implementation of Nonlinear Diffusion	145
9. OPTICAL FLOW	149
9.1 Introduction	149
9.2 Models for Optical Flow	150
9.2.1 A Differential Flow Model	151
9.2.2 Illumination Changes	153
9.2.3 Nonlinear Regularization	154
9.3 Analytical and Numerical Investigation	156
9.3.1 Consistency of the Model	156
9.3.2 Implementation and Numerics	160
9.4 Segmentation Methods	164
9.5 Experimental Results	166
9.5.1 Testing illumination changes	167
9.5.2 Recovery of motion boundaries	173
9.5.3 Standard test images	173
9.6 Discussion and Related Work	182
10. CONCLUSION	185
10.1 Summary	185
10.2 Outlook	188
Appendices	190
A Little Dictionary Of Wavelet Filters	191
Function Spaces	197
References	201

List of Figures

2.1	Various pavings of the time-frequency plane (κ, ω) .	17
3.1	A separable wavelet decomposition of the <i>Lena</i> test image.	33
3.2	The quincunx sampling lattice.	34
3.3	The HAAR scaling function indicator sets.	35
3.4	One step of a quincunx wavelet transform with the dilation matrix $\mathbf{Q}_{\text{quin},2}$. After filtering with the scaling and wavelet filter, the original image is downsampled by a factor of $\sqrt{2}$ and rotated by 45° leading to the two smaller <i>copies</i> containing averages (scaling function) and details (wavelet).	36
3.5	Hexagonal sampling lattice with seven channel sampling.	39
3.6	The hexagonal HAAR scaling function and the six associated wavelets.	40
3.7	The rhombic dodecahedron and the associated tessellation of \mathbb{R}^3 .	43
3.8	Geometrical interpretation of the sets C_n and D_n .	46
3.9	Upper and lower bounds for the areal coincidence ζ_n in the case of dyadic n -dimensional sampling.	47
4.1	Spatial resolution of wavelet frames (separable vs. non-separable case)	57
4.2	Illustration of Lemma 4.2.	64
4.3	The scaling function for Example 4.	66
4.4	The scaling function for Example 5.	69
4.5	The first orthogonal \mathcal{C}^1 scaling function for the quincunx dilation matrix $\mathbf{Q}_{\text{quin},2}$.	71
6.1	Absolute values of the angular errors.	108
6.2	Typical sparsity patterns for matrices associated to an operator in the standard form.	112

6.3	Typical sparsity pattern for matrices associated to an operator in the non-standard form.	114
8.1	Comparison between linear scale spaces.	137
8.2	PERONA-MALIK specific diffusivity , flux and first flux derivative.	139
8.3	SCHNÖRRS diffusivity function, flux and first flux derivative.	140
8.4	Comparison of linear and nonlinear diffusion methods.	141
8.5	Results of nonlinear wavelet estimation.	144
8.6	Results of nonlinear wavelet denoising.	145
8.7	Schematic illustration of wavelet-based AOS.	147
9.1	The aperture problem.	152
9.2	Input data for experiment 1.	168
9.3	Results of experiment 1.	168
9.4	Input data for experiment 2.	169
9.5	Results of experiment 2.	170
9.6	Comparison of illumination profiles for experiment 2.	170
9.7	Input data for experiment 3.	171
9.8	Results of experiment 3.	171
9.9	Comparison of illumination profiles for experiment 3.	172
9.10	Results of experiment 4.	173
9.11	Focus of motion boundary.	174
9.12	First and last frame of the <i>Office</i> test sequence.	174
9.13	Estimated optical flow for the <i>Office</i> test sequence.	175
9.14	Real optical flow for the <i>Office</i> test sequence.	175
9.15	First and last frame of the <i>Street</i> test sequence.	176
9.16	Estimated optical flow for the <i>Street</i> test sequence.	177
9.17	Real optical flow for the <i>Street</i> test sequence.	177
9.18	First and last frame of the RUBICs <i>cube</i> sequence.	178
9.19	Optical flow estimated for the RUBICs <i>cube</i> sequence.	179
9.20	First and last frame of the <i>Yosemite</i> sequence.	179
9.21	Estimated optical flow for the <i>Yosemite</i> test sequence.	180
9.22	Real optical flow for the <i>Yosemite</i> test sequence.	180
9.23	Estimated illumination profile for the <i>Yosemite</i> sequence.	181
A.1	Frame functions h , g^1 , g^3 and g^5 from the previous example.	195

List of Tables

3.1	Isotropy measures for generalized quincunx sampling in higher dimensions.	45
9.1	Average performance results for experiment 1.	169
9.2	Average performance results for experiment 2.	170
9.3	Average performance results for experiment 3.	172
9.4	Various flow results for the <i>Street</i> sequence.	178
9.5	Various flow results for the <i>Yosemite</i> sequence.	181
9.6	Impact of proposed model extensions.	182

Notation

The best notation is no notation.

—PAUL R. HALMOS

This thesis is mainly concerned with the processing of multidimensional data via wavelet methods. In order to make the text better readable, some specific multidimensional notation is collected first.

Throughout the whole text, operators and matrices are represented by bold-faced capital letters, i.e. $\mathbf{A}, \mathbf{B}, \dots \in \mathbb{R}^{n \times n}$, while vectors are given in the common notation, that is

$$\vec{x} = [x_1, x_2, \dots, x_n]^T \in \mathbb{R}^n,$$

where $[\dots]^T$ denotes the usual transposition. The power function of two vectors $\vec{x}, \vec{m} \in \mathbb{R}^n$ is defined by

$$\vec{x}^{\vec{m}} = [x_1^{m_1}, x_2^{m_2}, \dots, x_n^{m_n}]^T.$$

For $\vec{k} \in \mathbb{N}_0^n$ and a function $f(\vec{x})$ with $f : \mathbb{R}^n \rightarrow \mathbb{R}$, the vector differential operator will be written as

$$\mathbf{D}_{\vec{k}}(f)(\vec{x}) = \frac{\partial^{|\vec{k}|}}{(\partial \vec{x})^{\vec{k}}} f(\vec{x}), \quad \text{with } |\vec{k}| = k_1 + k_2 + \dots + k_n.$$

Still assuming $\vec{k} \in \mathbb{N}_0^n$, the vector faculty is expressed as

$$\vec{k}! = k_1! \cdot k_2! \cdot \dots \cdot k_n!.$$

With this statement at hand, the vector faculty function may be extended to non-integer and negative vector entries under usage of the EULER- Γ -function.

Finally, for $\vec{j} \in \mathbb{Z}^n$, the vector KRONECKER delta function is given by

$$\delta_{\vec{j}} = \begin{cases} 1 & \vec{j} = \vec{0} \\ 0 & \text{otherwise.} \end{cases}$$

Chapter 1

OUVERTURE

Jedem Anfang wohnt ein Zauber inne, der uns beschützt und der uns hilft zu leben

—HERMANN HESSE

The goal of the research presented in this thesis is to give an introduction to multidimensional non-tensor-product wavelet theory and its application to computer vision problems, especially those arising in the context of image and image sequence analysis.

Instead of directly delving into the theoretical work, the main topics of this thesis shall be introduced from a certain point of view first. We start by giving a brief historical overview of the development of wavelet theory within the last century from a pure mathematical concept to a widely applicable tool for diverse scientific problems. On one hand, this may just stimulate the readers interest to wavelet theory in general, but on the other hand, this survey shall also be viewed as an introduction to better understand the close relationships between the mathematical theory and its scientific relevance in physics, engineering, medicine, . . . (this enumeration is far from being exhaustive and could be prolonged arbitrarily). Because — varying a word of EINSTEIN — even the most beautiful mathematical theory is completely useless without any connection to the *real world*; and wavelets have in fact a very high significance for several concrete research areas.

Among these areas is also the research in computer vision, which shares the fundamental idea of the *scale concept* with the mathematical wavelet theory. This conceptual parallelism will give us a mathematical bridge leading to multiscale representations of images, theoretical results and several efficient

algorithms for computer vision problems that will be developed in the course of this thesis. To introduce the involved problems, the second part of this *overture* is dedicated to give an overview about what computer vision is and what its fundamental concepts are. In particular, we will describe some of its main roots, its connection to other scientific branches, which are very widespread and range from psychology and biology to information theory and (of course) applied mathematics. Furthermore, the general ideas, mathematical methods and problems are introduced to equip the reader with a basic feeling and understanding for the computer vision applications examined in the third part of this thesis.

This *overture* closes with a brief outline of the text explaining its structural organization and introducing the single chapters.

1.1. A Brief History of Wavelets

... the development of wavelets is an example where ideas from many different fields combined to merge into a whole that is more than the sum of its parts.

—INGRID DAUBECHIES

Within the last decade, wavelet theory has developed into an imposing mathematical tool, that finds application in a big variety of mathematical as well as engineering, physical, ... problems. However, things started about a century ago with the fundamental work of ALFRED HAAR [101].¹ His famous construction of the function system

$$\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k) \quad \text{with} \quad \psi(x) = \begin{cases} 1 & \text{if } x \in [0, 1/2), \\ -1 & \text{if } x \in [1/2, 1), \\ 0 & \text{else} \end{cases}$$

gave rise to the first orthonormal wavelet basis of the function space $L_2(\mathbb{R})$, even though the term *wavelet* was not invented for a long time yet. While HAARs original results were primarily of interest for pure mathematicians, most of the fundamental concepts of wavelet theory originate from his work, since HAARs construction gave the first time-frequency representation for functions $f \in L_2(\mathbb{R})$, which means a unification of two classical representations of functions: Of course, the arithmetical notion of a function as a varying entity associated to another entity, which may be interpreted as a time- or space-dependent relation is in fact much older and goes back to ideas of JOHANN BERNOULLI (1718) and was formulated in full mathematical strength by KARL

¹An even much older mathematical concept that obeys the principles of wavelet representations is musical notation, whose roots are found in Egypt and date back to about 2700 B.C. On the other hand, today's notation system, which is the first that implicitly contains *wavelet ideas* was invented in the Florentine school during the *basso continuo period* around the year 1600.

WEIERSTRASS about 150 years later. The idea of splitting functions into parts of different frequencies was first introduced by JEAN BAPTISTE FOURIER in 1807 in the context of investigations of heat diffusion processes [87].

The first conceptual generalizations of the HAAR system in terms of time-frequency representations of functions originated about 1925, these were the introduction of WALSH function series [229], which may be seen as the first prototype of a *wavelet packet basis* [52] on one hand and the functions nowadays called FRANKLIN wavelets [88] [111], which belong to the class of piecewise linear spline wavelets on the other hand. Afterwards, at the latest from 1930 on, the development of theory started to grow in parallel within many different disciplines such as harmonic analysis, engineering or quantum mechanics (often independently and without awareness of similar results in other scientific branches) and therefore, it would expand the scope of this introduction to consider all the aspects and influences that led to today's wavelet theory — this might require (and in fact deserves by the authors' opinion) a whole book on its own. For this reason, we will concentrate on some few but striking events in the theories' evolution that followed the abovementioned pioneering works.

A major contribution from communications engineering is the SHANNON sampling theorem (1949)², which states, that every frequency-banded function, that is sampled at least with its NYQUIST rate, may be reconstructed exactly by the sampled values. This theorem, which caused a little revolution in radioelectronics and neighbouring disciplines must be viewed in the greater context of parallel developments in harmonic analysis, where distributional FOURIER theories came up while a group of mathematicians around WIENER, ZYGMUND, PALEY, HARDY and LITTLEWOOD [139] [178] [177] were building up the foundations of a *generalized harmonic analysis*. One result of particular interest in the context of wavelets is the so-called LITTLEWOOD-PALEY decomposition, which yields a dyadic decomposition of a given function in the FOURIER domain, from whose atoms, the LITTLEWOOD-PALEY blocks, one may derive simple L_p characterizations for functions in more general function spaces. Such characterizations are direct predecessors of BESOV space characterizations by thresholded wavelet coefficients, which on the other hand play an important role in optimal lossy coding, denoising or nonlinear diffusion processes in computer vision [73] [41] (this connection will become clearer in Part III of this thesis). Later on COIFMAN, WEISS [51] and others were trying to find atomic expan-

²The sampling theorem is named after SHANNON, nevertheless it was stated and proven independently of SHANNON and some years before him by WHITTAKER (in 1929) and also by KOTELNIKOV (1933). The sampling theorem originally goes back to CAUCHY's *cardinal series theorem* (1841) and to ideas of HADAMARD and DE LA VALLÉE-POUSSIN concerning the theory of entire functions.

sions of functions and general characterizations of certain function spaces, like SOBOLEV or HARDY spaces (see also Appendix B), which also play a fundamental role in the theory of partial differential equations. These works finally led to STRÖMBERG's construction [215] of orthonormal spline wavelet bases for HARDY spaces in 1983.

However, let's turn back to the past for a moment: Inspired by the discovery of the HEISENBERG uncertainty principle (1927), the physicist and engineer DENNIS GABOR gave a new time-frequency decomposition of functions in terms of the *windowed FOURIER transform* [89] in order to overcome the limitations in spatial processing of the classical FOURIER transform. Similar ideas led to further contributions like the *ZAK transform* or the *ambiguity transform* and to the formulation of the *frame concept* for nonharmonic series [79]. Though all these inventions gave better time-frequency decompositions than the classical methods, they were still unsatisfying in various aspects, for example, GABOR's transform yielded a uniform time-frequency zoom into the (x, ω) -plane.³ This was overcome by the introduction of scale, the next remarkable development in the step towards wavelet theory, namely the discovery of CALDERÓN's *reproducing formula* [35]

$$f(x) = \int_{\mathbb{R}_+} \int_{\mathbb{R}} \int_{\mathbb{R}} \psi(s^{-1}, v - x) \cdot f(w) \cdot \psi(s^{-1}, v - w) \frac{dv dw ds}{s} \quad (1.1)$$

for a well-suited distribution ψ . This is exactly today's well-known continuous wavelet transform formula; the reader familiar with wavelet theory should note the similarity between the inner integral and the wavelet coefficients on one scale.⁴ Although the identity (1.1) dates back to 1964, it took 20 more years since its discrete version was developed and the full meaning of this transform was realized. . .

Meanwhile, engineers working on signal and image processing problems, were trying to find sets of linear filters (filter banks) that obey a perfect reconstruction property, which is desirable for several communication and coding applications. This finally led to the discovery of *quadrature mirror filters* [55] in 1976; these type of filters do not only yield unaliased subband coding schemes

³Another aspect concerning the non-optimality that comes along with the GABOR transform is the celebrated BALIAN-LOW theorem [13] [141], which states that a frame consisting of any two-parameter family of window functions $f_{m,n} = e^{-2\pi i m x} \cdot f(x - n)$ with $m, n \in \mathbb{Z}$ is either spatially or frequency locally completely unlocalized, i.e. either $\|f_{m,n}\|^2$ or $\|\hat{f}_{m,n}\|^2$ have an infinite second moment.

⁴Seen from group theory, the windowed FOURIER transform and the wavelet transform are the same [99]: While the first is the square-integrable representation of the WEYL-HEISENBERG group of shifts and phase-shifts, the latter one is the accordant representation of the affine group. Recently, a unification of these concepts via a certain intertwining LIE group operator could be achieved [119] [203].

but are also a major component in the multiresolution analysis setting applied to generate orthonormal wavelet bases. A short time later, BURT and ADELSON formulated the LAPLACE-pyramid as an alternative method for efficient image coding [29]; from today's point of view, this concept belongs into the class of (non-tight) wavelet frames. Related ideas led DESLAURIERS and DUBUC to the *dyadic interpolation* scheme [72]. The moment, wavelets became the popular tool they are nowadays coincides with the moment of understanding, that all the mathematical and physical theories and the engineering methods could be formulated in a unifying manner and efficient algorithms are accessible simultaneously. This breakthrough was achieved by STÉPHANE MALLAT and YVES MEYER in autumn of 1986 and was published in the fundamental article [144], where also the first discrete variant of the reproducing formula (1.1) appeared. Even a few years earlier, G. RICHTER applied discrete HAAR transforms independently in the context of astronomical image analysis [186] [187]. Consequently, his construction must be seen as the *real* first occurrence of the discrete wavelet transform, even though it was not built in the same generality as MALLAT's method and also did not achieve the same popularity.

After this large step ahead was once made, things evolved in a breathtaking speed: In 1988, INGRID DAUBECHIES constructed the first compactly supported orthonormal wavelet bases of arbitrary regularity [60] with the consequence, that cut-off errors could be completely avoided. Wavelets became more and more attractive for various applications, when some limitations of orthonormal wavelets — for example non-symmetry and boundary handling problems — could be overcome by the introduction of *biorthogonal* wavelets [49] [225], *wavelet frames* [61] and *wavelets on intervals* [50]. The mathematical foundations for wavelet based numerical analysis were built by BEYLKIN, COIFMAN and ROKHLIN in [20], while the treatment of PDEs was first addressed in [92]. The first attempts to non-trivial multidimensional wavelet generalizations were made by KOVAČEVIĆ and VETTERLI, who is also one of the pioneers of subband coding theories, for the orthogonal case [127] and by COHEN and DAUBECHIES for the biorthogonal case [47], but the achieved results were not very satisfying, since either worse regularity or extremely huge filters (thousands of taps) had to be accepted as drawbacks. Some improvements in this direction could be realized by RON and SHEN for higher dimensional tight wavelet frames via box-spline constructions [195].

The period of rapid developments lasted about seven or eight years, in which nearly the whole theory was built up. Within the last few years the barycenter of wavelet research tended more and more to concrete applications rather than new theoretical results. Many of these applications deal with the processing and evaluation of higher dimensional data (e.g. images, atmospheric data,

medical data, ...), but on the other hand, most of the related algorithms rely on intrinsically onedimensional wavelet methods due to the abovementioned disadvantages coming along with multidimensional wavelet systems. But usually, the objects or structures in higher dimensional signals are not reducible to (essentially) onedimensional tensor product entities and from this point of view, it seems to be more natural to consider real multidimensional wavelets as bases (or frames). The overcoming of this gap is one of the fundamental intentions of this thesis: In the upcoming investigations, we will describe new ideas how to create and apply wavelets in more than one dimension (Part I), we will generalize several principles of onedimensional wavelet analysis to the more general case (Part II) and finally, we will consider concrete algorithms for the developed concepts in computer vision applications (Part III).

1.2. About Computer Vision

Das Bild ist ein Modell der Wirklichkeit.

—LUDWIG WITTGENSTEIN

Computer vision is the theory of processing and manipulating discrete (or better: discretized) visual information by a machine. It is concerned with the task of finding *meaningful* and *useful* information within the given visual patterns of data. Of course, the definitions of meaningfulness and usefulness considered in this context highly depend on the target of the processing and on the *complexity level* of abstraction.⁵ Typical high-level applications are e.g. to provide robots with the capability of *seeing* in order to employ them in industrial manufacturing or to implement autonomous driving and navigation systems into vehicles of any kind. Applications like these can be subdivided into a set of physical models of the environment and a set of lower-level tasks like image representation, image segmentation, object recognition, regularization, ... *acting on* the physical models. Due to the fact that one deals with visual data (images), *image processing* is a major component in nearly all computer vision problems. The partitioning of a problem into a set of subproblems is a very natural technique to simplify the situation on one hand, but it also supplies one with the possibility to incorporate the vision problem into an even bigger context: from (visual) information to a *learning* process which shall endow

⁵Many scientists in computer vision divide vision tasks into three layers of abstraction, these are *low-level*, *intermediate-level* and *high-level* vision tasks. Roughly, low-level vision means the computational theory standing at the beginning of any vision tasks. The intermediate-level deals with representation of information and the transformations to model the processes, while all the *higher* physical and model-theoretic conforming laws are collected under the term high-level vision. Only a well-formulated high-level vision layer makes sensible physical realizations of a vision system in a concrete environment feasible at all [153]. Despite this separation in terminology, in many practically oriented problems the differences between the three layers are usually rather diffuse than sharp.

the machine with some *competence* in order to *act* by own means. This is the essential problem of implementing a perception-action cycle — a principal goal of many cognitive scientists. Formally, this is described by the relational scheme

$$\text{syntax} \longleftrightarrow \text{semantics} \longleftrightarrow \text{pragmatics} .$$

The visual system of humans (and also that of most animals) has highly evolved capabilities. Consequently, a natural approach to computer vision is that of *simulating* the human (or animal) visual system on a machine. This approach involves biological and psychological knowledge about visual perception in order to *reproduce* the events occurring in the retinal ganglion cells and in the brain. For example, the existence of neurophysiological and psychological evidences for a processing on multiple scales in *biological vision* systems [243] [249] gives rise to embed computer vision processes into a corresponding mathematical multiscale framework as it is given for example by wavelet theory. On the retinal level, this corresponds to choose a certain mathematical way of *representing* the visual information. On the other hand, the *cortical processing* should be mapped onto a set of *information processing* tasks like segmentation or extraction of boundaries, motions, . . . , depending on the application in mind. The cortical level is usually formulated mathematically by differential equations, variational problems or stochastic modelling to connect the represented visual data with the physical assumptions (kinematics, illumination models, . . .) about the environment in order to make reliable conclusions about observable *events* within the perceived information. Anyway, the reader should keep in mind that while a biological lifeform always acts autonomous, the machine unconditionally requires human interaction (since the human *supervisor* defines the goals of the machine at the beginning and is also the final evaluating institution by all means).

Most higher level problems in vision are (at least partially) mathematically *ill-posed* in the sense of HADAMARD. This fact has various causes, one very fundamental one goes back to the fact that visual perception involves a *reduction* of information from the threedimensional world to a twodimensional image plane. Other problems making vision processes ill-posed (at least on the computational level) are immediate changes of exterior conditions (e.g. somebody is switching on a light), the finiteness of optical resolution or the incorporation of noise. There are strong indications that humans (as well as animals) are able to deal with these problems by accumulated *experience*, that is, biological vision is closely tied to the actions of the individuum and its attending and responding to dynamic changes in the perceived environment. These observations have lead to the *active vision* methodology, see e.g. [12]. On the other hand, there must be some fundamental mechanisms to solve basic ill-posed problems implemented in biological vision systems, since the external stimuli leading

to experiences must be processed somehow by the individuum. Therefore, in computer vision, the efficient mathematical solution of such problems is a very fundamental task and the problems we will discuss in Part III of this thesis will essentially be posed and solved as such. Hereby, the wavelet formalism will again turn out to be a strong and efficient tool.

1.3. Organization of the Thesis

This work is divided into three parts. The first part treats the theory, construction principles and requirements to build multidimensional wavelets. It starts with a presentation of the connections between wavelets, FOURIER theories and harmonic analysis and recalls the fundamentals of onedimensional wavelet theory to equip the reader with the basic mathematical concepts and some ideas about the difficulties in higher dimensions. In the following chapter, the properties of matrix subsampling in higher dimensions are investigated in some detail. This forms the fundamental for multidimensional multiresolution analyses, the main concept to build nonseparable wavelets in higher dimensions. The main chapter of Part I is the final one, it addresses the issue of designing multidimensional wavelet filters. Hereby, several design aspects like regularity, vanishing moments, linear phase property, ... are considered and weighed against each other. The advantages and disadvantages of different wavelet families (orthogonal, biorthogonal, frames) are discussed and some new and much simpler construction methods are also derived.

Part II is concerned with wavelet analysis in higher dimensions. It begins with a generalization of several 1D wavelet tools like the MALLAT algorithm, properties of projection operators and the statement of various useful approximation results. This introductory chapter is followed by a survey of the representation of operators in wavelet bases. The concept of *connection coefficients*, including their calculation is transferred into the multidimensional case and some relationships to discrete operators are derived. This chapter is closed by the discussion of the two principle ways to represent operators in terms of wavelet bases. Afterwards, the solution of ill-posed problems using wavelet methods is investigated. This concluding chapter of the second part builds the bridge to the intended computer vision applications, since these latter ones usually occur as mathematically ill-posed problems. Several different numerical treatments of such problems are considered and optimized for operator equations described in terms of multidimensional wavelets.

The final part of this thesis deals with computer vision applications, mainly with the task of the analysis of images and image sequences. First, linear and nonlinear *scale-spaces*, their description by PDEs and integral equations, as well as their connections to wavelet representations of images and certain func-

tion spaces are discussed. Especially the involved characterizations of nonlinear scale-spaces will be a helpful tool in the problem of regularizing vision tasks depending on noisy data. In the following, the main computer vision problem considered in this thesis shall be introduced and solved: the computation of the optical flow for a given sequence of images. All the theoretical and practical results obtained so far will flow into the approach(es) we propose and apply. Various flow models and different additional phenomena as illumination changes or occlusions are taken into account. The presented results will be compared and discussed in relation to the outcome of various other successful approaches that appeared so far.

The thesis is closed by some concluding remarks and a brief outlook into remaining problems, further generalizations and other thinkable applications. Finally, two appendices provide the interested reader with a collection of wavelet and scaling filters that are (at least indirectly) used in this work and with a survey about the function spaces the considered problems were mathematically embedded into.

I

MULTIDIMENSIONAL WAVELETS

Chapter 2

WARMING UP: ONEDIMENSIONAL WAVELETS

A journey of a thousand miles must begin with a single step.

— LAO-TSE

To start with the presentation of the mathematical concepts of multidimensional wavelets, it is quite useful to consider the onedimensional case first. This usefulness stems from two main facts: First, the presentation of the simpler onedimensional theory allows one to carry over some of the main results that do not depend on the dimensional structure of the considered physical space directly to higher dimensions and simultaneously, the problems that arise for other theorems get directly evident. The second point is, that it is also possible to derive *multidimensional* wavelets directly from their onedimensional *prototypes* via tensor product constructions. These wavelets are separable in the spatial domain and yield some principal disadvantages, however, they build the *natural bridge* between one- and multidimensional wavelet bases; walking over this bridge is the best way to nonseparable multidimensional wavelets.

This chapter is subdivided into three sections. In the first section, the basic paradigms of (not necessarily onedimensional) wavelet theory are derived from the FOURIER transform and its properties. Starting from the uncertainty principle of communication, the GABOR transform and later on the introduction of scale leading to the wavelet transform are motivated. The notion of *multiresolution analysis* is introduced here and the special meaning of dyadic dilations is also enlightened. Section 2 mainly deals with the conditions to build onedimensional wavelet bases via multiresolution analyses. Several fundamental assertions concerning the design of wavelet filters for one dimension are presented. Additionally, we discuss some special properties of wavelets which are

desirable for certain applications. Finally, the third section classifies the presented results into a context of related works. The presentations in the whole chapter are mostly very brief and partially not held in full generality. This is due to its introductory character and shall not be seen as incompleteness but as a reduction to the essential; the multidimensional case will of course be discussed in a much wider extent.

2.1. Wavelet Paradigms — From the Fourier Transform to Multiresolution Analyses

The representation of a signal as a function of time does not bring out the frequencies in play, whereas on the other hand, the Fourier representation conceals the moment of emission and the duration of each of the components of the signal. An adequate representation ought to combine the advantages of the two complementary descriptions; it ought also to be in a discrete form, which is better suited to the theory of communications.

—ROGER BALIAN

The FOURIER transform is the classical tool to decompose a function f into parts of different frequencies. While it exactly specifies the allotments of all oscillations of f , no information about the local spatial behaviour of f is accessible in the FOURIER domain. This is due to a very well-known mathematical principle.

THEOREM 2.1 (UNCERTAINTY PRINCIPLE OF COMMUNICATION, 1946)
Suppose that $f \in L_2(\mathbb{R})$ with $\|f\|_{L_2} = 1$. Then

$$u_f(\xi, \omega) = \int_{\mathbb{R}} (x - \xi)^2 \cdot |f(x)|^2 dx \cdot \int_{\mathbb{R}} (\omega - \omega)^2 \cdot |\hat{f}(\omega)|^2 d\omega \geq \frac{1}{4} \quad (2.1)$$

holds for all $\xi, \omega \in \mathbb{R}$.

Seen from the FOURIER transform point of view, we may interpret the uncertainty principle of communication as follows. The basis functions in the frequency domain are the 2π -periodic functions $f_k(x) = e^{ikx}$, $k \in \mathbb{Z}$ for which one easily verifies

$$\begin{aligned} \int_{\mathbb{R}} (x - \xi)^2 \cdot |f_k(x)|^2 dx &= \infty & \text{and} \\ \int_{\mathbb{R}} (\omega - \omega)^2 \cdot |\hat{f}_k(\omega)|^2 d\omega &= 2\pi \cdot (k - \omega)^2. \end{aligned}$$

This confirms the abovementioned statement about the bad spatial resolution capabilities of the FOURIER transform and moreover, one is left with an infinite uncertainty u_f except for the trivial case $f \equiv \text{const.}$ In order to overcome these

limitations, DENNIS GABOR suggested to consider a *windowed* version of the FOURIER transform [89].¹

DEFINITION 2.2 (WINDOWED FOURIER/GABOR TRANSFORM)

$$T_\psi(f)(x, \omega) = \int_{\mathbb{R}} f(\xi) \cdot \psi(\xi - x) \cdot e^{-i\omega\xi} d\xi.$$

Hereby, $\psi(x)$ is a (rapidly decaying) distribution often referred to as *window function*.² Probably the most popular choice for $\psi(x)$ is the normalized GAUSS function

$$\psi(x) = \sqrt{\frac{s}{\pi}} \cdot e^{-sx^2}, \quad s > 0,$$

since this function satisfies the uncertainty relation (2.1) with the optimal value $u_\psi \equiv \frac{1}{4}$ and gives thus an ideally accurate time-frequency localization for functions $f \in L_2(\mathbb{R})$. But on the other hand, one is also interested to have a *good representation* in the sense, that the family of functions $\psi(x - n) \cdot e^{-2\pi i m x}$ forms an (orthonormal) basis or a frame for $L_2(\mathbb{R})$ in order to be also able to get a stable reproduction of the original function from a (in some sense) *simplified* phase-space version. Unfortunately, several years later it was found that this hope can never be fulfilled for any kind of GABOR frame transform no matter how the window function $\psi(x)$ is chosen. This is the main statement of the famous BALIAN-LOW theorem.

THEOREM 2.3 (BALIAN-LOW, 1981/1985) *Suppose that $f \in L_2(\mathbb{R})$ and*

$$f_{m,n}(x) = e^{2\pi i m x} \cdot f(x - n), \quad m, n \in \mathbb{Z}.$$

If $\{f_{m,n} : m, n \in \mathbb{Z}\}$ is a frame for $L_2(\mathbb{R})$ then either

$$\int_{\mathbb{R}} x^2 |f(x)|^2 dx = \infty \quad \text{or} \quad \int_{\mathbb{R}} \omega^2 |\hat{f}(\omega)|^2 d\omega = \infty.$$

Proof. See [62], pp. 108–112. □.

¹The uncertainty principle of communication possesses many parallel statements in somewhat related scientific areas. Maybe the most famous among these is the HEISENBERG uncertainty principle of quantum mechanics. It mainly states that the impulse and the spatial localization of any elementary particle cannot be exactly determined at the same time, see also [175].

²In practice, some additional conditions on the distribution $\psi(x)$ are sensible: In order to obtain sharp time-frequency localization for the GABOR transform, one might require $\psi(\cdot)$ to be in the SCHWARTZ class $\mathcal{S}(\mathbb{R})$ of rapidly decaying functions. A weaker possible condition would be

$$\text{id} \cdot \psi(\cdot) \in L_2(\mathbb{R}) \quad \text{and} \quad \text{id} \cdot \hat{\psi}(\cdot) \in L_2(\mathbb{R}).$$

Remark. If one oversteps the critical NYQUIST sampling rate, that means one considers a family

$$f_{m,n}(x) = e^{i\omega_0 m x} \cdot f(x - x_0 n), \quad m, n \in \mathbb{Z}$$

with $\omega_0 \cdot x_0 < 2\pi$, it is possible to generate (even tight) frames with a good phase-space localization. However, orthogonal windowed FOURIER bases with this property can never exist.

The fundamental idea of the wavelet transform is now to replace the *phase-shift* maintained by the factor $e^{2\pi i m x}$ by a *scaling* $x = x_0 \cdot s^{-j}$ of the spatial variable. Hereby, $|s| > 1$ denotes the *scaling* (or *dilation*) *factor* and $j \in \mathbb{Z}$ is the *scale number*. In the discrete case, it is sensible to choose x_0 to be the coarsest sampling level and to require $j \in \mathbb{N}_0$. These are the prerequisites for the following definition.

DEFINITION 2.4 (DISCRETE WAVELET TRANSFORM) *Given a function $\psi \in L_2(\mathbb{R})$ such that*

$$c_\psi = 2\pi \cdot \int_{\mathbb{R}} \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty \quad (2.2)$$

and consider the two-parameter family

$$\psi_{j,k}(x) = |s|^{-j/2} \cdot \psi(s^{-j}x - k), \quad k \in \mathbb{Z}, j \in \mathbb{N}_0,$$

with $s \in \mathbb{Z} \setminus \{-1, 0, 1\}$. The discrete wavelet transform of a function $f \in L_2(\mathbb{R})$ is then defined to be³

$$W_\psi(f)(j, k) = \int_{\mathbb{R}} f(\xi) \cdot \psi_{j,k}(\xi) d\xi. \quad (2.3)$$

A function ψ that satisfies (2.2) is called wavelet.

Remark. The *admissibility condition* (2.2) guarantees the stable invertibility of the wavelet transform (2.3). This technical requirement is mild and shall not be discussed here in further detail. We only point out that the wavelet transform

³The *continuous wavelet transform* can be analogously defined by

$$W_\psi(f)(s, x) = \sqrt{s/c_\psi} \cdot \int_{\mathbb{R}} f(\xi) \cdot \psi\left(\frac{\xi - x}{s}\right) d\xi, \quad s, x \in \mathbb{R}.$$

This definition is more general than its discrete counterpart and yields a highly redundant description of the function. Since this thesis targets at applications that require discretizations at some stage of processing, only the discrete version is considered here. By introducing the theory of wavelet frames later in this thesis, we will obtain another way to get redundant representations, which might be desirable for some applications as well as for numerical stabilization. The reader interested in the continuous wavelet transform may look into the books [147] or [140].

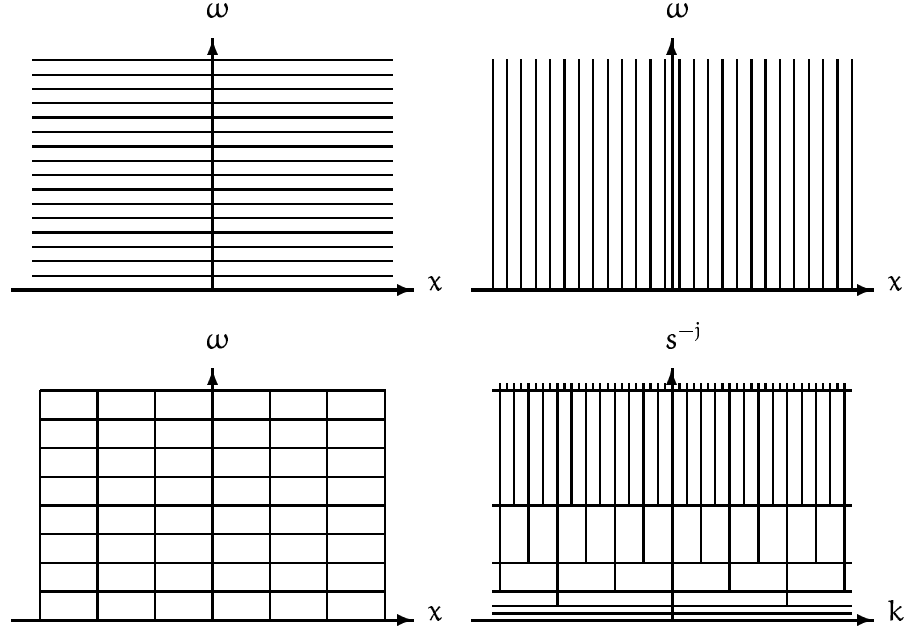


Figure 2.1. Various pavings of the time-frequency plane (x, ω) . *Upper left*: FOURIER domain. *Upper right*: DIRAC pulses. *Lower left*: GABOR transform. *Lower right*: Wavelet basis representation (note that $k \sim x$ and $s^{-j} \sim \omega$).

is an isometry and its inverse is thus given by the adjoint operator W_{ψ}^* .

By construction, the wavelet transform leads indeed to a stable and well localized time-scale representation of $L_2(\mathbb{R})$ functions. Moreover, it can be shown, that similar statements are possible for functions in several other spaces like $L_p(\mathbb{R})$ with $1 < p < \infty$ or certain SOBOLEV or BESOV spaces. However, the question that arises at this point is what type of functions ψ are *good* wavelets and how to construct them. From the mathematical point of view, a good idea would be to have a function family $\psi_{j,k}$ that forms an orthonormal basis of $L_2(\mathbb{R})$. This directly leads to the concept of multiresolution analyses, which will be introduced now.

DEFINITION 2.5 (MULTIRESOLUTION ANALYSIS) A multiresolution analysis (MRA) is a nested sequence of function spaces $V_j \subset L_2(\mathbb{R})$, $j \in \mathbb{Z}$, that

fulfills the following conditions for some $s \in \mathbb{Z}$:

$$V_{j+1} \subset V_j \quad \text{for all } j \in \mathbb{Z}, \quad (2.4)$$

$$f(x) \in V_j \quad \text{if and only if } f(s^j \cdot x) \in V_0 \quad \text{for all } j \in \mathbb{Z}, \quad (2.5)$$

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\}, \quad (2.6)$$

$$\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L_2(\mathbb{R}), \quad (2.7)$$

$$\exists \varphi(x) \in V_0 \quad \text{such that } \{ \varphi(x - k), k \in \mathbb{Z} \} \text{ is an orthonormal basis for } V_0. \quad (2.8)$$

The generator function $\varphi(x)$ is called *scaling function* or *father wavelet* of the MRA. We additionally remark, that the condition (2.8) may be relaxed by assuming that $\{ \varphi(x - k), k \in \mathbb{Z} \}$ forms a RIESZ basis for V_0 .⁴

The conditions (2.4) and (2.5) have the intuitive meaning that V_j is $|s|$ times as big as V_{j+1} and hence the orthogonal complement W_{j+1} of V_{j+1} in V_j is spanned by $|s| - 1$ functions. Seen from the communications engineering point of view, $|s|$ is the number of different *channels*, the signal is passed through. The wavelet transform itself may thus be interpreted as a bandpass filtering of a signal. Anyway, the case $s = \pm 2$ is of special interest, because only one function is needed to span the orthogonal complement; the function generating this orthogonal complement $\psi(x) \in W_0$ will then be denoted as *mother wavelet* or just *wavelet*. The desirability of such *dyadic dilations* will become clear in the upcoming section, when concrete multiresolution analyses will be created.

2.2. Construction of Onedimensional Wavelets

So far we have described what the (discrete) wavelet transform is and we introduced the concept of MRAs, which are the fundamentals in designing orthonormal wavelet bases. Now, we will care about the concrete design of scaling functions and wavelets. We start by giving a general representation for scaling functions that mainly goes back to the property (2.5).

⁴Recall that a family $\{ \varphi(x - k), k \in \mathbb{Z} \}$ forms a RIESZ basis for

$$V = \overline{\text{span}\{ \varphi(x - k), k \in \mathbb{Z} \}}$$

if and only if there exist constants $C_u > C_l > 0$ such that for every sequence $\{\gamma_j\}$, $\gamma_j \in \mathbb{R}$ we have

$$C_l \cdot \left(\sum_k |\gamma_k|^2 \right)^{1/2} \leq \left\| \sum_k \gamma_k \varphi(\cdot - k) \right\| \leq C_u \cdot \left(\sum_k |\gamma_k|^2 \right)^{1/2}.$$

LEMMA 2.6 *Every scaling function of a MRA satisfies a scaling equation, that is, there exists a sequence $\{h_k\}_{k \in \mathbb{Z}}$ of real numbers such that*

$$\varphi(x) = |s|^{1/2} \cdot \sum_{k \in \mathbb{Z}} h_k \cdot \varphi(s \cdot x - k). \quad (2.9)$$

Proof. It follows from Definition 2.5 that $\varphi(x) \in V_{-1}$ if and only if $\varphi(s \cdot x) \in V_0$. Together with (2.8) it follows, that φ must be written in a linear combination of the form (2.9). \square .

The scaling equation thus relates the scaling function to a *discrete filter* $[h_k]_{k \in \mathbb{Z}}$, which explains — in combination with the hierarchical scaling structure — the close connection between MRA wavelets and subband coding schemes. Taking the FOURIER transform of (2.9), we directly obtain

$$\hat{\varphi}(\omega) = H(\omega/s) \cdot \hat{\varphi}(\omega/s). \quad (2.10)$$

The new function $H(\omega)$ is called *transfer function* or *symbol* of the filter $[h_k]_{k \in \mathbb{Z}}$. It is verified to be given by

$$H(\omega) = |s|^{-1/2} \cdot \sum_{k \in \mathbb{Z}} h_k \cdot e^{-ik\omega}. \quad (2.11)$$

The symbol has the advantage, that it is analytically more accessible than the scaling equation and many properties of the scaling function and the wavelet may be expressed via $H(\omega)$ as we shall see. Equipped with this formulation, we are at first ready to give a characterization of the desired pairwise orthonormality of the shifts $\varphi(\cdot - k)$.

PROPOSITION 2.7 *Let $\varphi(x) \in L_2(\mathbb{R})$ be a scaling function as in (2.9). Then the following statements are equivalent:*

$$\langle \varphi(x), \varphi(x - k) \rangle_{L_2} = \delta_k. \quad (2.12)$$

$$\sum_{k \in \mathbb{Z}} h_k \cdot h_{k+sm} = \delta_m. \quad (2.13)$$

$$\sum_{j=0}^{|s|-1} |H(\omega + 2\pi j/|s|)|^2 \equiv 1. \quad (2.14)$$

Proof. See [184], pp. 95–102. \square .

Remark. The reader should carefully note, that not every scaling function, that fulfills (2.13) or (2.14) automatically generates an orthonormal function system in $L_2(\mathbb{R})$. On the other hand, there exist various mild criterions on the

symbol (or the filter) that directly guarantee the L_2 existence of φ .⁵ We point out without (the very simple) proofs, that all the scaling functions presented in this thesis satisfy such criteria.

From now on, we only consider dyadic dilations for the rest of this section, that means we assume $s = 2$. As mentioned before, this case is especially desirable, since only one function (wavelet) is needed to span the orthogonal complement in a subspace V_j and thus, its scaled and translated versions make up the complete MRA. Moreover, the wavelet for a dyadic dilation is very easy obtainable.

LEMMA 2.8 *Given a scaling function $\varphi(x)$ whose integer translates $\varphi_k(x) = \varphi(x - k)$, $k \in \mathbb{Z}$ form an orthogonal series spanning a subspace $V_0 \subset L_2(\mathbb{R})$. Then its orthogonal complement W_0 of $V_0 \subset V_{-1}$ is spanned by the integer translates of the function*

$$\psi(x) = \sqrt{2} \cdot \sum_{k \in \mathbb{Z}} (-1)^k \cdot h_{1-k} \cdot \varphi(2 \cdot x - k).$$

*Moreover, $\{\psi_{j,k}(\cdot) = 2^{-j/2} \psi(2^{-j} \cdot - k) \mid j, k \in \mathbb{Z}\}$ forms an orthonormal basis for $L_2(\mathbb{R})$.*⁶

⁵Two widely known criteria on the symbol $H(\omega)$ satisfying the relation (2.14) are the following ones [44] [134]:

CRITERION (A. COHEN) *Let $H(0) = 1$ and a set K congruent to $[-\pi, \pi] \bmod 2\pi$ be given such that*

- i. $\int_{j>0} \inf_{\omega \in K} |H(s^{-j}\omega)| > 0$
- ii. K contains a neighbourhood of $\omega = 0$.

Then the scaling function $\varphi(x)$ generates an orthonormal sequence in $L_2(\mathbb{R})$.

CRITERION (W. LAWTON) *Suppose without loss of generality, that the filter h_k is supported in $[0, n]$. Let $H(0) = 1$ and define a $2n-1 \times 2n-1$ matrix \mathbf{A} by*

$$(a_{j,k}) = \sum_{m=0}^n h_m h_{k-sj+m}, \quad 1-n \leq j, k \leq n-1.$$

If the eigenvalue 1 of \mathbf{A} is nondegenerate, then the associated scaling function $\varphi(x)$ generates an orthonormal sequence in $L_2(\mathbb{R})$.

It can be shown that the criteria of COHEN and LAWTON are equivalent. They mainly rely on convergence properties of the *cascade algorithm*, that is the repeated application of (2.10) leading to

$$\hat{\varphi}(\omega) = \hat{\varphi}(0) \cdot \prod_{m>0} H(s^{-m}\omega). \quad (2.15)$$

⁶For non-dyadic dilations, finding the wavelet filters becomes significantly harder — this is especially true for the multidimensional case. Formally, the problem reduces to complete a given vector containing the

Proof. The orthogonality $\varphi_k \perp \psi$ can be directly verified by simply calculating the inner product and using the pairwise orthogonality of the φ_k . The last assertion follows from the construction of a MRA via (2.5) and (2.8). \square .

The reader may verify, that the function ψ satisfies the admissibility condition with $c_\psi = 2 \ln 2$ and is thus indeed a wavelet in the sense of Definition 2.4.

As a next point, we may ask, what kind of wavelet properties are *good* for certain applications and how to achieve them. From the mathematical point of view, it seems natural to ask for wavelets, which have a number of vanishing moments, i.e. all polynomials up to a certain order, say n , are already orthogonal to $\psi_{j,k}(x)$ for all $j, k \in \mathbb{Z}$. This property may be characterized as follows.

LEMMA 2.9 *Let φ be a scaling function associated with a filter $[h_k]_{k \in \mathbb{Z}}$ and ψ be the related wavelet. Then the following are equivalent:*

$$\int_{\mathbb{R}} x^m \cdot \psi_{j,k}(x) dx = 0 \quad j, k \in \mathbb{Z}, \quad \forall m \leq n, \quad (2.16)$$

$$\frac{d^m}{d\omega^m} H(\omega) \Big|_{\omega=\pi} = 0 \quad \forall m \leq n. \quad (2.17)$$

$$\sum_{k \in \mathbb{Z}} (-1)^k \cdot k^m \cdot h_k = 0 \quad \forall m \leq n. \quad (2.18)$$

Proof. See e.g. [140]. \square .

Relations of type (2.16) are called *moment cancellations*, while (2.17) is usually called *STRANG-FIX condition* (after [85]) and (2.18) is mainly referred to as *sum rules up to order m*. Equipped with this, we are ready to attack the problem of designing filters that lead to onedimensional compactly supported wavelets. Since orthogonality as well as vanishing moments may be described in terms of the symbol $H(\omega)$, the starting point for this will be the relation (2.14). Several methods to obtain filters from this equation exist:

- 1 DAUBECHIES' original construction [60]. Suppose the wavelet shall have m vanishing moments. This may be characterized by (2.17) and yields a factorization

$$|H(\omega)|^2 = H(\omega) \overline{H(\omega)} = \cos^{2m} \left(\frac{\omega}{2} \right) \cdot |p(\omega)|^2, \quad (2.19)$$

polyphase components of the scaling filter to a paraunitary matrix. This matrix can always be easily obtained by a GRAM-SCHMIDT procedure [96], however, the resulting wavelets are no longer finitely supported and may suffer from worse spatial resolution. The derivation of *compactly supported wavelets* from a given scaling function is a generally unsolved problem, it was attacked by several wavelet researchers, see e.g. [189] [136]. A certain way to obtain such wavelets for a special case is presented in Chapter 3.2.2.

with some real trigonometric polynomial $p(\omega)$. Now using the shorthands $y = \cos^2(\omega/2)$ and $P(1 - y) = |p(\omega)|^2$, we can rewrite (2.14) to

$$y^m \cdot P(1 - y) + (1 - y)^m \cdot P(y) \equiv 1$$

with $P \geq 0$. This equation is solved by the BEZOUT polynomials

$$P(y) = \sum_{k=0}^{m-1} \binom{m-1+k}{k} \cdot y^k + y^m \cdot Q(y),$$

where $Q(y)$ is an arbitrary odd polynomial.

2 The second approach is due to MEYER [157]. Here, one considers

$$|H(\omega)|^2 = 1 - \lambda_m \cdot \int_0^\omega \sin^{2m-1}(t) dt$$

and has to choose the factor $\lambda_m = \frac{\Gamma(m+1/2)}{\Gamma(1/2)\Gamma(m)}$ to guarantee $H(\pi) = 0$.

3 Maybe the most elegant way is to expand the trigonometric identity

$$\left(\sin^2(\omega) + \cos^2(\omega) \right)^{2m-1} \equiv 1$$

and to rearrange the terms in a special way to obtain an equation that fits (2.14). This was done in [214].

These methods — as well as some others not mentioned here — lead to a description of the squared symbol $|H(\omega)|^2$. In order to eliminate the filter coefficients h_k from this representation, we have to take the square root of this expression, which is maintained by the following classical result.

THEOREM 2.10 (FEJER-RIESZ) *Let $f(\omega)$ be a nonnegative trigonometric cosine polynomial with real coefficients,*

$$f(\omega) = \sum_{k=0}^n a_k \cos(k\omega), \quad a_k \in \mathbb{R}.$$

Then there exists a trigonometric polynomial $F(\omega) = \sum_{k=0}^n b_k e^{ik\omega}$ with real coefficients b_k such that

$$|F(\omega)|^2 = f(\omega).$$

The classical proof is found in [190], a constructive one yielding all coefficients b_k is given in [184]. \square .

At this stage it is urgently induced to give a first anticipation of the general multidimensional case.

LEMMA 2.11 *The FEJER-RIESZ Theorem 2.10 does not generalize to higher dimensions.*

A counterexample with detailed explanation can be found in [198]. \square .

Lemma 2.11 is the main obstacle in general multidimensional wavelet constructions. The absence of factorization theorems causes all abovementioned filter design methods to fail here. One way to circumvent this problem is the usage of *tensor product* wavelets, which can be directly built up from the onedimensional *prototypes*; these constructions only lead to spatially separable wavelets — a property which is undesired especially if one works with visual data. Another more general way around the factorization problem can be direct design methods, a certain one of these will be introduced (for the onedimensional case) in the sequel. The main idea of this new concept is the decomposition of wavelet filters into a set of *basic filters*.

DEFINITION* 2.12 *Denote by $[h_k]^{*m}$ the m times repeated convolution of the discrete filter $[h_k]_{k \in I}$, i.e.*

$$[h_k]^{*m} = \underbrace{[h_k]_{k \in I} * [h_k]_{k \in I} * \cdots * [h_k]_{k \in I}}_{m \text{ times}}.$$

As next, we introduce the candidates for our *basic filters* itself, which are defined via

$$[g_k^{l,m}] = [1 \quad -1]^{*l} * [1 \quad 1]^{*m}. \quad (2.20)$$

These filters have been known as *binomial filters* in the engineering society for a while. The building HAAR filters $[1 \quad 1]$ and $[1 \quad -1]$ are also referred to as *fundamental filters*, since all basic filters may be obtained from these. The basic filters have the following useful features.

LEMMA* 2.13 *The filter $[h_k^m] = [g_k^{0,m}]$ satisfies the sum rules up to order $m - 1$.*

Proof. Applying a classical theorem due to PASCAL, we can directly verify the representation

$$[h_k^m] = \left[\binom{m}{0} \binom{m}{1} \cdots \binom{m}{m} \right].$$

Inserting this into (2.18) and using the induction principle, the result follows together with the binomial identity $\binom{m+1}{k+1} \equiv \binom{m}{k} + \binom{m}{k+1}$. \square .
An immediate consequence of Lemma 2.13 is

COROLLARY* 2.14 All filters $[g_k^{l,m}]$ satisfy the sum rules up to order $m - 1$.

Obviously, any filter $[g_k^{l,m}]$ is of length $l + m + 1$ and without loss of generality, its coefficients may be numbered by $g_0^{l,m}, g_1^{l,m}, \dots, g_{l+m}^{l,m}$. Now, we can prove the following nice property of a certain collection of basic filters.

PROPOSITION* 2.15 Let \mathbf{A}_n be the $(n + 1) \times (n + 1)$ matrix

$$\mathbf{A}_n = \begin{bmatrix} h_0^n & g_0^{1,n-1} & g_0^{2,n-2} & \dots & g_0^{n,0} \\ h_1^n & g_1^{1,n-1} & g_1^{2,n-2} & \dots & g_1^{n,0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_n^n & g_n^{1,n-1} & g_n^{2,n-2} & \dots & g_n^{n,0} \end{bmatrix},$$

then

$$\det \mathbf{A}_n = (-2)^{\frac{n(n+1)}{2}}.$$

Especially, $\det \mathbf{A}_n \neq 0$ for all $n \in \mathbb{N}_*$ and from this, we directly deduce that the $n + 1$ filters

$$\{h^n, g^{1,n-1}, g^{2,n-2}, \dots, g^{n,0}\}$$

form a linear independent family of basic filters and moreover, every subfamily

$$\{h^n, g^{1,n-1}, g^{2,n-2}, \dots, g^{n-k,k}\} \quad (2.21)$$

additionally satisfies all sum rules up to order $k - 1$ by Lemma 2.13 and Corollary 2.14. Furthermore, this family of basic filters yields a natural decomposition of every scaling filter that satisfies the sum rules up to order $k - 1$ into an even (symmetric) and an odd (antisymmetric) part since h^n is always even and $g^{j,n-j}$ is even for j even and odd for j odd.

Proof of Proposition 2.15. Again, we will make use of the induction principle. For $n = 1$ we easily obtain

$$\mathbf{A}_1 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{and} \quad \det \mathbf{A}_1 = -2.$$

In the second step, we will evaluate $\det \mathbf{A}_{n+1}$ from $\det \mathbf{A}_n$. In particular, we acquire

$$\begin{aligned}
\det \mathbf{A}_{n+1} &= \begin{vmatrix} h_0^{n+1} & g_0^{1,n} & g_0^{2,n-1} & \dots & g_0^{n+1,0} \\ h_1^{n+1} & g_1^{1,n} & g_1^{2,n-1} & \dots & g_1^{n+1,0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ h_{n+1}^{n+1} & g_{n+1}^{1,n} & g_{n+1}^{2,n-1} & \dots & g_{n+1}^{n+1,0} \end{vmatrix} \\
&= \sum_{j=0}^{n+1} (-1)^{n+1+j} \cdot g_j^{n+1,0} \cdot \begin{vmatrix} \vdots & \vdots & \vdots & \dots & \vdots \\ h_i^{n+1} & g_i^{1,n} & g_i^{2,n-1} & \dots & g_i^{n,1} \\ \vdots & \vdots & \vdots & \dots & \vdots \end{vmatrix}_{i \neq j} \\
&= (-1)^{n+1} \cdot \sum_{j=0}^{n+1} \binom{n+1}{j} \cdot \begin{vmatrix} \vdots & \vdots & \vdots & \dots & \vdots \\ h_i^n + h_{i+1}^n & \dots & g_i^{n,1} + g_{i+1}^{n,1} \\ \vdots & \dots & \vdots \end{vmatrix}_{i \neq j} \\
&= (-1)^{n+1} \cdot \sum_{j=0}^{n+1} \binom{n+1}{j} \cdot \det \mathbf{A}_n \\
&= (-2)^{n+1} \cdot \det \mathbf{A}_n.
\end{aligned}$$

By induction, we obtain the desired relation

$$\det \mathbf{A}_n = \det \mathbf{A}_1 \cdot \prod_{j=2}^n (-2)^j = (-2)^{\frac{n(n+1)}{2}}. \quad \square.$$

It is thus possible to describe all onedimensional wavelet filters by unique linear combinations built over filter families of the type (2.21). For a better understanding, we shall give another interpretation of the abovemade construction. The B-spline $B_1(x)$ of zeroth order is identical to the HAAR scaling function, which may also be described by the discrete filter $[1 \ 1] = h^1$. In this sense, the scaling function associated to a wavelet with m vanishing moments may thus be also specified by the spatial convolution of the B-spline $B_{m+1}(x)$ with a well-suited distribution $r(x)$ in order to achieve orthogonality. This convolution product is given via (2.21) in the filter coefficient domain. We close this section by a short example, that shall explain the filter design procedure.

EXAMPLE* 1 We will describe all orthogonal scaling filters of length 8 yielding at least 2 vanishing moments. Following Proposition 2.15, the space of these filters is a subspace of the span

$$\{h^7, g^{1,6}, g^{2,5}, g^{3,4}, g^{4,3}\}.$$

Establishing orthogonality via (2.13) or (2.14), we obtain the filters

$$\frac{1}{128} \cdot h^7 + k \cdot g^{1,6} + (64 \cdot k^2 - \frac{7}{256}) \cdot g^{2,5} + \frac{\gamma}{2} \cdot g^{3,4} + (64 \cdot \gamma + \frac{35}{1024} + 224 \cdot k^2 - 262144 \cdot k^4) \cdot g^{4,3}$$

where k is a free parameter and γ satisfies

$$\gamma = \frac{k}{2^9} \cdot \frac{-63 \cdot 2^7 \cdot k - 5 \cdot 2^{22} \cdot k^3 + 2^{35} \cdot k^5 - \sqrt{385 \cdot 2^{16} \cdot k^2 - 91 \cdot 2^{29} \cdot k^4 + 9 \cdot 2^{44} \cdot k^6 - 2^{56} \cdot k^8 - 945}}{1 + 2^{14} \cdot k^2}.$$

The variable k may for example be used to optimize the regularity or symmetry properties of the according scaling function or to maximize the energy compaction in the low pass band of the filter. Using the GRÖBNER basis algorithm, closed form descriptions for higher order filter coefficients are also accessible.

2.3. Related Work

Many of the work related to the material in this chapter was already mentioned in the short wavelet history at the beginning of this text. For sake of completeness, we recall the main contributions here. The classical FOURIER transform was first formulated in 1807 and published in 1822 in FOURIER'S main work [87]. The theory of the windowed FOURIER transform goes back to [89], while frames were introduced in [79]. A more recent article on the theory of GABOR functions and frames containing several new results is [116].

As already mentioned, the continuous wavelet transform originally goes back to CALDERÓN [35], who introduced the first reproducing integral transform based on a time-scale decomposition. This idea was rediscovered in the middle of the 80's by several mathematicians of the new french harmonic analysis school (Y. MEYER, P.G. LEMARIÉ, G. BATTLE) — in [137] the term *wavelet* was first used in the context of reproducing kernels. Anyway, a similar construction was done about five years earlier by STRÖMBERG [215] and therefore, Y. MEYER calls him “the inventor of wavelets”.

The main contributors to the discrete wavelet theory are S. MALLAT, who was the first to realize the relations between subband filtering schemes, reproducing integral transforms and discrete time-scale decompositions (see [144]) and I. DAUBECHIES who is responsible for the construction of compactly supported wavelets [60], which just caused the popularity, wavelet transforms achieved within the last decade. Several other constructions followed, e.g. [224], [225] or the direct method presented at the end of this chapter, which is due to the author [169]. A similar method to this one, but incorporating less design features was independently introduced in [1].

Chapter Summary

We have introduced the most fundamental principles of FOURIER analysis. From these, we have derived the concept of scaling and the notion of a multiresolution analysis, leading to the basic ideas of discrete wavelet theory. Furthermore, the general characterization of onedimensional wavelets in terms

of the related discrete filter were presented. Out of these relations, a new method for onedimensional filter design that does not rely on spectral factorizations (and is thus extendable to multiple dimensions) could be developed. Finally, a short example illustrating the working way of this new method was given.

Chapter 3

MULTIDIMENSIONAL SIGNAL PROCESSING — MRA AND SAMPLING IN \mathbb{R}^n

It should be stated at the outset place that the multivariable theory is much less developed and much more complicated than the one-variable theory.

—PAVEL WOJTASZCZYK

The computation, evaluation and efficient storing of higher dimensional data — like images or image sequences in the case of computer vision tasks — naturally requires multidimensional signal processing theories. In the discrete case, such a theory mainly relies on the concepts of linear filtering and affine sampling, since these are involved in the discretization of several basic mathematical operations like function interpolation, partial differentiation or dilation. Moreover, they also build the fundamental for wavelet transforms in arbitrary dimensions. This is the point of consideration in the actual chapter.

In Section 1, we will generalize the concept of a multiresolution analysis (MRA) to the multidimensional case and we will introduce the related scaling equation and the new *scaling matrix*, which mainly describes the up- and down-sampling between various scales within the multidimensional MRA. In Section 2, the most common scaling matrices and their induced sampling lattices in the twodimensional case — which is of special interest for the applications regarded in this thesis — are investigated and some simple transformation examples are given. Section 3 deals with the more general n -dimensional case, which only plays a minor role in this work. Nevertheless, we will deduce a general design principle for scaling matrices and give the first classification of a certain family of matrices yielding dyadic sampling. This section is closed by some new results regarding the sampling quality in terms of an isotropy measure and its asymptotical behaviour for increasing numbers of dimensions.

Finally, some possible extensions and some other articles related to the contents of this chapter are quoted.

3.1. Multiresolution Analyses in Higher Dimensions

This section is mainly devoted to the multidimensional generalization of the MRA setting developed in Chapter 2. The reformulation of the basic definitions and properties is straightforward:

DEFINITION 3.1 (MULTIDIMENSIONAL MULTIREOLUTION ANALYSIS) *A multidimensional multiresolution analysis for $L_2(\mathbb{R}^n)$ is a sequence of nested subspaces $V_j \subset L_2(\mathbb{R}^n)$, $j \in \mathbb{Z}$, that fulfills the following conditions for some expansive matrix \mathbf{Q} :*

$$\mathbf{Q}(\mathbb{Z}^n) \subset \mathbb{Z}^n \quad \text{i.e.} \quad \mathbf{Q} \in \text{Mat}(n \times n, \mathbb{Z}) \quad (3.1)$$

$$V_{j+1} \subset V_j \quad \text{for all } j \in \mathbb{Z}, \quad (3.2)$$

$$f(\vec{x}) \in V_j \quad \text{if and only if } f(\mathbf{Q}^j \cdot \vec{x}) \in V_0 \quad \text{for all } j \in \mathbb{Z}, \quad (3.3)$$

$$\bigcap_{j \in \mathbb{Z}} V_j = \{0\}, \quad (3.4)$$

$$\overline{\bigcup_{j \in \mathbb{Z}} V_j} = L_2(\mathbb{R}^n), \quad (3.5)$$

$$\exists \varphi(\vec{x}) \in V_0 \quad \text{such that } \{ \varphi(\vec{x} - \vec{k}), \vec{k} \in \mathbb{Z}^n \} \quad \text{is an orthonormal basis for } V_0. \quad (3.6)$$

\mathbf{Q} is called scaling matrix or dilation matrix. Hereby, the matrix \mathbf{Q} is said to be expansive, if all eigenvalues $\lambda_{i,\mathbf{Q}}$ of \mathbf{Q} are strictly larger than 1, which has the geometrical interpretation, that all principal axes of the linear transform induced by \mathbf{Q} are stretched, i.e. \mathbf{Q} evokes genuine sub-sampling in all directions.

From this definition, we may directly deduce the multidimensional equivalent to the scaling equation formulated in Lemma 2.6:

COROLLARY 3.2 *Every scaling function of a n -dimensional MRA satisfies a scaling equation. There exists a multisequence $\{h_{\vec{k}}\}_{\vec{k} \in \mathbb{Z}^n}$ of real numbers such that*

$$\varphi(\vec{x}) = |\det \mathbf{Q}|^{1/2} \cdot \sum_{\vec{k} \in \mathbb{Z}^n} h_{\vec{k}} \cdot \varphi(\mathbf{Q} \cdot \vec{x} - \vec{k}). \quad (3.7)$$

Proof. Similar to that of Lemma 2.6. □.

Again, the scaling equation may be transformed into the FOURIER domain. This directly leads to

$$\hat{\phi}(\vec{\omega}) = H(\mathbf{Q}^{-T}\vec{\omega}) \cdot \hat{\phi}(\mathbf{Q}^{-T}\vec{\omega}), \quad (3.8)$$

where \mathbf{Q}^{-T} denotes the transposed inverse matrix of \mathbf{Q} . The *transfer function* or *symbol* $H(\vec{\omega})$ can be evaluated to

$$H(\vec{\omega}) = |\det \mathbf{Q}|^{-1/2} \cdot \sum_{\vec{k} \in \mathbb{Z}^n} h_{\vec{k}} \cdot e^{-i \cdot \langle \vec{k}, \vec{\omega} \rangle}. \quad (3.9)$$

From the algebraic fact, that the number of different cosets of $\mathbf{Q}^T \cdot \mathbb{Z}^n$ in \mathbb{Z}^n is equal to $|\det \mathbf{Q}|$, another immediate consequence for the MRA can be formulated.

COROLLARY 3.3 *In accordance to the remarks following Definition 2.5, a multidimensional MRA with an associated scaling matrix \mathbf{Q} with $\det \mathbf{Q} = q$ requires $|q| - 1$ wavelets ψ^i , $i = 1, \dots, |q| - 1$ to be complete.*

The sub-sampling evoked by \mathbf{Q} keeps every q -th sample and discards the rest, which means that the local information of q points is packed into one point on the next scale. By means of the cosets of $\mathbf{Q}^T \cdot \mathbb{Z}^n$ in \mathbb{Z}^n , we define the VORONOI cell of the dilation matrix \mathbf{Q} to be the set

$$A_{\mathbf{Q}} = \left\{ \vec{x} \in \mathbb{R}^n \mid \|\vec{x}\| \leq \|\vec{x} - \vec{p}\|, \forall \vec{p} \in \mathbf{Q}^T \cdot \mathbb{Z}^n \setminus \{0\} \right\},$$

that describes the local area of information concentration of the sub-sampling. This is all we require for the moment. Equipped with the concept of scaling matrices, we shall closer investigate their properties and influence on the wavelet transform next. We mainly focus on the twodimensional case, since this shows all properties of multidimensional filtering in a quite vivid manner and additionally this instance also plays the major role in image processing.

3.2. Uniform Affine Sampling Lattices in \mathbb{R}^2

What is more beautiful than the Quincunx, which, from whatever direction you look, is the same.

—MARCUS FABIVS QUINTILIAN

The scaling matrix \mathbf{Q} is the higher dimensional pendant to the scaling factor s in the previous chapter. However, due to the increased number of spatial dimensions, we have to take much more care in choosing \mathbf{Q} than choosing s , which was (more or less) straightforward. Following Corollary 3.3, it would again be a good idea to maintain a scaling matrix \mathbf{Q} with $\det \mathbf{Q} = \pm 2$. But

additionally, it also makes sense to have a nearly *isotropic* sub-sampling, which means that all directions of the data grid should be sampled in a very similar way. Full isotropy would correspond to radial symmetric sampling, which is of course impossible since a set of even-sized circles can never form a disjoint covering of the plane. As an approximation of radial symmetric sampling, one might use convex regular tessellations of the considered space \mathbb{R}^n , that means using a scaling matrix \mathbf{Q} such that its VORONOI cell is a convex regular polytope with a space tiling property. This shall be investigated for the case $n = 2$ in the following. We start with the remark, that there exist exactly three different convex and regular tessellations of \mathbb{R}^2 , these are given by the equilateral triangle, the square and the regular hexagon.

3.2.1. Regular Sampling

The simplest regular and convex tessellation of \mathbb{R}^2 is that consisting of squares. A corresponding sampling may be realized by extending both axes with the same factor ν . In order to keep the number of associated wavelets as small as possible, $\nu = 2$ is the optimal and very widely used choice. This yields the scaling matrix

$$\mathbf{Q}_{\text{reg}} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}. \quad (3.10)$$

One directly verifies, that 3 wavelets are necessary to build up a MRA for \mathbf{Q}_{reg} . This choice of scaling matrix is also implicitly used in tensor product wavelets built from a dyadic onedimensional wavelet basis: Given a scaling function and a wavelet $\varphi(x), \psi(x) \in L_2(\mathbb{R})$, one can directly build tensor product functions by

$$\begin{aligned} \phi(x, y) &= \varphi(x) \cdot \varphi(y), & \psi^1(x, y) &= \varphi(x) \cdot \psi(y), \\ \psi^2(x, y) &= \psi(x) \cdot \varphi(y), & \psi^3(x, y) &= \psi(x) \cdot \psi(y). \end{aligned}$$

Then, the function family

$$\left\{ \psi_{j, \vec{k}}^i(\cdot) = |\det \mathbf{Q}|^{-j/2} \psi^i(\mathbf{Q}^{-j} \cdot -\vec{k}) \mid j \in \mathbb{Z}, \vec{k} \in \mathbb{Z}^2, i \in \{1, 2, 3\} \right\}$$

forms an orthonormal basis for $L_2(\mathbb{R}^2)$. The related filters are obviously given by simple convolutions of the onedimensional filters. Such wavelets have become very popular for several applications, since they are easy to build (as seen above) and additionally, the wavelet transform itself is also easy to implement, since it can be subdivided into onedimensional transforms. It is not exaggerated to assert, that such separable wavelets are used in more than 95% of all higher dimensional wavelet applications. The working way of transforms with dilation matrix \mathbf{Q}_{reg} is illustrated in Figure 3.1.



Figure 3.1. Two steps of a separable wavelet decomposition of the *Lena* test image. In each transformation step, the (remaining) image is decomposed into low-low-pass (remaining local averages $W_\phi f$, cf. (2.3)), high-low-pass (vertical details $W_{\psi_1} f$), low-high-pass (horizontal details $W_{\psi_2} f$) and high-high-pass (diagonal details $W_{\psi_3} f$) information channels. Most of the significant information is contained in the low-low-pass channel. The sampling factor of 2 along each spatial direction cf. (3.10) causes the halving of the axes lengths in each transformation step.

The original motivation for this regular sub-sampling was to achieve a sampling pattern close to a circular VORONOI cell (radial symmetry). In order to assess the *quality* of this sampling scheme, we evaluate the *areal coincidence* ζ_{reg} between the circle A_{circ} and the induced square A_{reg} of the same area as a *sampling isotropy measure*. It is defined by

$$\zeta_{\text{reg}} = \frac{\text{vol}(A_{\text{reg}} \cap A_{\text{circ}})}{\text{vol}(A_{\text{circ}})}.$$

The value $\zeta_{\text{reg}} = 0.9094 \dots$ is easily verified, which means that the regular sampling induced by \mathbf{Q}_{reg} achieves about 90.94% of a perfect radially symmetric sampling for functions with an ideal uniform spectrum.

3.2.2. The Quincunx Lattice

The regular sampling achieves good sampling quality, but on the other hand we have $\det \mathbf{Q}_{\text{reg}} = 4$ which is not optimal and also leads to much more complicated schemes for building the wavelets from the scaling function in the case of nonseparable filters (see also the last endnote of Chapter 2). A regular two-dimensional two-channel sampling can be realized by an axis scaling of $\sqrt{2}$ and a rotation about $\pi/4$ (or a similar reflection) to *stay on the grid*. This so called *quincunx sampling* is maintained by both of the following scaling matrices:

$$\mathbf{Q}_{\text{quin},1} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \quad \text{or} \quad \mathbf{Q}_{\text{quin},2} = \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

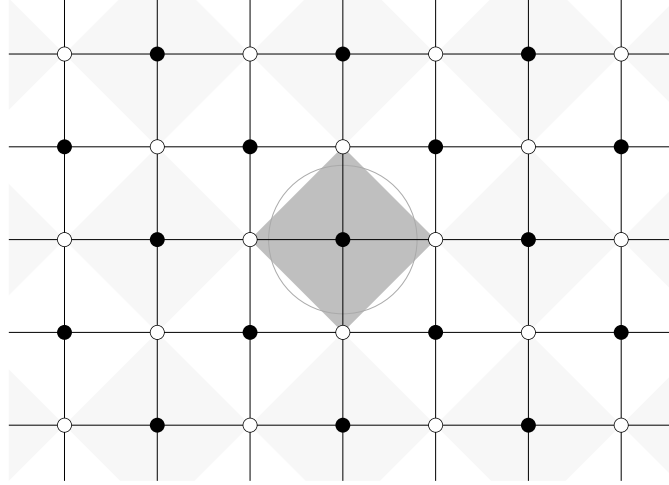


Figure 3.2. The quincunx sampling lattice. — The shaded area is the VORONOI cell A_{quin} of information concentration, while the thin circle indicates how *ideal sampling* would look like. The black and white dots represent the two cosets of $\mathbf{Q}_{\text{quin}} \cdot \mathbb{Z}^2$ in the sampling lattice.

These dilation matrices induce a VORONOI tessellation with $\pi/4$ rotated squares of area 2, this is shown in Figure 3.2. Such dilations are good candidates for various vision purposes, because the human eye is less sensitive to resolution along diagonals than horizontals or verticals [127] that appear in regular sampling schemes. Since the circle is rotation invariant, the sampling quality again achieves 90.94% of the ideal value in terms of the areal coincidence. Although the scaling matrices $\mathbf{Q}_{\text{quin},1}$ and $\mathbf{Q}_{\text{quin},2}$ look very similar and generate the same sampling lattice, they imply remarkable differences, which will be demonstrated now. Since the matrix $\mathbf{Q}_{\text{quin},1}$ fulfills the identity $\mathbf{Q}_{\text{quin},1}^2 = 2 \cdot \mathbf{Id}$, iterated application of equation (3.8) leads to the spectral factorization

$$\begin{aligned} \hat{\varphi}(\vec{\omega}) &= \prod_{k=1}^{\infty} H\left((\mathbf{Q}^{-T})^k \vec{\omega}\right) = \prod_{k=1}^{\infty} H(2^{-k} \vec{\omega}) \cdot \prod_{k=1}^{\infty} H(2^{-k} \mathbf{Q}^{-T} \vec{\omega}) \\ &= \hat{\varphi}(\omega_1, \omega_2) \cdot \hat{\varphi}(\omega_1/2 + \omega_2/2, \omega_1/2 - \omega_2/2), \end{aligned}$$

which also implies a spatial factorization of the scaling function $\varphi(\vec{x})$.¹ This *factorability* is explained by the fact, that $\mathbf{Q}_{\text{quin},1}$ is similar to the antidiagonal

¹If one uses onedimensional filters with the scaling matrix $\mathbf{Q}_{\text{quin},1}$, this leads directly to *separable* wavelets, since then the symbol $H(\omega)$ is also onedimensional. This utilization of such plain filters gives a very simple construction of multidimensional wavelets of arbitrary smoothness, since the regularity properties can be

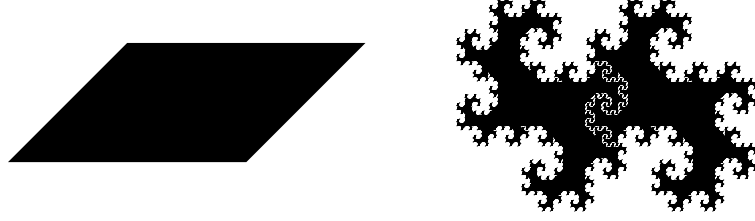


Figure 3.3. The HAAR scaling function indicator sets for $\mathbf{Q}_{\text{quin},1}$ (left) and $\mathbf{Q}_{\text{quin},2}$ (right).

matrix $\begin{bmatrix} 0 & 1 \\ 2 & 0 \end{bmatrix}$, which incorporates a directed dilation followed by a change of axes. In comparison to this, the second quincunx dilation matrix $\mathbf{Q}_{\text{quin},2}$ has complex-valued eigenvalues $\lambda_{\pm} = 1 \pm i$ inhibiting such similarities. We also have $\mathbf{Q}_{\text{quin},2}^2 \neq 2 \cdot \mathbf{Id}$ but $\mathbf{Q}_{\text{quin},2}^8 = 16 \cdot \mathbf{Id}$, which makes a decomposition as above impossible and always leads to nonseparable wavelets. To illustrate this fundamental difference between the scaling matrices, we show the indicator sets χ_{φ} of the associated HAAR scaling functions and wavelets in Figure 3.3. While the first matrix indeed yields a separable scaling function indicator set, namely a simple parallelogram, the second matrix creates a fractal called *twin dragon* as the indicator set of the related HAAR scaling function.

A step of a quincunx wavelet transform looks like Figure 3.4 shows. There occurs one obvious problem: Given some discrete data on a rectangular grid and applying a wavelet transform step to this data, one obtains two rotated and downsampled data sets which cannot be directly represented on a computer. The data has to be reorganized into a rectangular structure before the processing can proceed. There are various ways to do this, two of them will be explained now. First, one could rewrite the downsampled data set as follows

$$\begin{vmatrix} x & o^1 & x & o^2 \\ o^3 & x & o^4 & x \\ x & o^5 & x & o^6 \\ o^7 & x & o^8 & x \end{vmatrix} \longrightarrow \begin{vmatrix} o^1 & o^2 & o^6 \\ o^3 & o^4 & o^8 \\ o^7 & o^5 & \end{vmatrix},$$

to maintain the $\pi/4$ rotation. In this scheme, x denotes samples to be dropped by the quincunx downsampling. This type of rearrangement leads again to a

directly carried over from the onedimensional prototypes. A generalization of this approach to other scaling matrices \mathbf{Q} with $|\det \mathbf{Q}| = 2$ and $\mathbf{Q}^2 = \pm 2\mathbf{Id}$ such that all eigenvalues of \mathbf{Q} are real numbers can be found in [17]. This approach leads to nonseparable wavelets for $\mathbf{Q}_{\text{quin},1}$, but these have some structural deficiencies making them inappropriate for most applications. To mention the main drawback, the filters are aligned along two rows leading to a strongly favoured sampling along the row direction, which gives similar results than separable filtering. Moreover, this construction cannot yield *good* wavelets for the matrix $\mathbf{Q}_{\text{quin},2}$.



Figure 3.4. One step of a quincunx wavelet transform with the dilation matrix $\mathbf{Q}_{\text{quin}, 2}$. After filtering with the scaling and wavelet filter, the original image is downsampled by a factor of $\sqrt{2}$ and rotated by 45° leading to the two smaller *copies* containing averages (scaling function) and details (wavelet).

data set on a rectangular grid and must be repeated in every further transformation step. One drawback of this method is that storage space is wasted in the corners of such an arrangement, since there occur (many for big data sets) zero entries.

The second method to be presented relies on the idea of *wavelet packet decompositions* [52] [199]. This generalization of MRA gives a decomposition not only of the low pass but (at least possibly) also of the high pass part of a sampled signal.² The quincunx scheme can be kept on regular grids in this sense

²The original idea behind wavelet packet decompositions is that of the search for a *best basis* in terms of its coding entropy. Having a signal $f \in V_0$ existing on some level on a MRA and applying one wavelet transform step to f , we obtain a lowpass part $f_l \in V_{-1}$ and a highpass part $f_h \in W_{-1}$. In a standard wavelet transform, only f_l will be further transformed to reduce the coding size of f_l . But probably, an additional

by doing two transform steps in one, which corresponds to a rotation of the grid about $\pi/2$ or 0 — this keeps the grid structure by all means. However, if one does so, the high pass portion of data after the first step has also to be further transformed. Therefore, this *double quincunx* scheme gives similar to the regular case a four-channel sampling. Having scaling and wavelet equations of the form

$$\begin{aligned}\varphi(\vec{x}) &= |\det \mathbf{Q}|^{1/2} \cdot \sum_{\vec{k} \in \mathbb{Z}^n} h_{\vec{k}} \cdot \varphi(\mathbf{Q} \cdot \vec{x} - \vec{k}) \quad \text{and} \\ \psi(\vec{x}) &= |\det \mathbf{Q}|^{1/2} \cdot \sum_{\vec{k} \in \mathbb{Z}^n} g_{\vec{k}} \cdot \varphi(\mathbf{Q} \cdot \vec{x} - \vec{k})\end{aligned}$$

double application easily leads to

$$\begin{aligned}\varphi(\vec{x}) &= |\det \mathbf{Q}| \cdot \sum_{\vec{k} \in \mathbb{Z}^n} \sum_{\vec{l} \in \mathbb{Z}^n} h_{\vec{k}} \cdot h_{\vec{l}} \cdot \varphi(\mathbf{Q}^2 \cdot \vec{x} - \mathbf{Q} \cdot \vec{k} - \vec{l}) \\ &= |\det \mathbf{Q}| \cdot \sum_{\vec{m} \in \mathbb{Z}^n} \tilde{h}_{\vec{m}} \cdot \varphi(\mathbf{Q}^2 \cdot \vec{x} - \vec{m})\end{aligned}$$

with $\tilde{h}_{\vec{m}} = \sum_{\vec{k} \in \mathbb{Z}^n} h_{\vec{k}} \cdot h_{\vec{m}-\mathbf{Q}\vec{k}}$ as the new scaling filter. The wavelets $\psi^1(\vec{x})$, $\psi^2(\vec{x})$ and $\psi^3(\vec{x})$ correspond to three other filters, which are build analogously via

$$\tilde{g}_{\vec{m}}^1 = \sum_{\vec{k} \in \mathbb{Z}^n} h_{\vec{k}} \cdot g_{\vec{m}-\mathbf{Q}\vec{k}}, \quad (3.11)$$

$$\tilde{g}_{\vec{m}}^2 = \sum_{\vec{k} \in \mathbb{Z}^n} g_{\vec{k}} \cdot h_{\vec{m}-\mathbf{Q}\vec{k}} \quad \text{and} \quad (3.12)$$

$$\tilde{g}_{\vec{m}}^3 = \sum_{\vec{k} \in \mathbb{Z}^n} g_{\vec{k}} \cdot g_{\vec{m}-\mathbf{Q}\vec{k}}. \quad (3.13)$$

The double quincunx scheme is very similar to the regular sampling presented in Section 3.2.1, but there is one fundamental difference: While it is very difficult to derive compactly supported wavelets from a given scaling function on a regular grid, the double quincunx scheme gives such wavelets in a straightforward manner by the above relations (3.11)–(3.13). In the applications presented in Part III, a wavelet transform based upon this sampling scheme is employed successfully.

transform of f_h might also give a sparser representation of f_h . This is what the wavelet packet algorithm does: An incoming signal is completely decomposed and afterwards it is checked, which *decomposition trees* require the smallest amount of storage. Finally, this coding is applied.

This ends the discussion of quincunx sampling matrices. We close this section with an overview over all dyadic dilation matrices $\mathbf{Q} \in \text{Mat}(2 \times 2, \mathbb{Z})$ for \mathbb{R}^2 . In particular, one obtains the following classification:

PROPOSITION* 3.4 *In two dimensions, there exist — up to similarity transforms — exactly six different dyadic scaling matrices, these are the already-mentioned quincunx matrices $\mathbf{Q}_{\text{quin},1}$ and $\mathbf{Q}_{\text{quin},2}$ as well as*

$$\begin{bmatrix} -1 & -1 \\ 1 & -1 \end{bmatrix}, \quad \begin{bmatrix} 0 & -2 \\ 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} -1 & -2 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & -2 \\ 1 & 0 \end{bmatrix}.$$

Proof. A straightforward eigenvalue analysis and simple case distinctions give the result. \square .

All of these matrices also give either a quincunx subsampling or a horizontal/vertical subsampling lattice, since these are obviously the only possibilities for dyadic subsampling in two dimensions. But except from the first of these sampling matrices, which describes a rotation of $3\pi/4$, the latter three dilation matrices possess no desirable symmetry properties as the ones presented before and are thus used rather seldom in practice.

3.2.3. The Hexagonal Case

... eine endlose Erfindungslust in der Abwandlung und allerfeinsten Ausgestaltung eines und immer desselben Grundschemas, des gleichseitig-gleichwinkligen Sechsecks, herrschte da; aber in sich selbst war jedes der kalten Erzeugnisse von unbedingtem Ebenmaß und eisiger Regelmäßigkeit, ja, dies war das Unheimliche, Widerorganische und Lebensfeindliche daran; sie waren zu regelmäßig, die zum Leben geordnete Substanz war es niemals zu diesem Grade, dem Leben schauderte vor der genauen Richtigkeit, es empfand sie als tödlich, als das Geheimnis des Todes selbst, und Hans Casdorp glaubte zu verstehen, warum Tempelbaumeister der Vorzeit absichtlich und insgeheim kleine Abweichungen von der Symmetrie in ihren Säulenordnungen angebracht hatten.

—THOMAS MANN [150]

So far we only contemplated the grid $\mathbb{Z} \otimes \mathbb{Z}$, which is of course especially well-suited for computer processing purposes. Unfortunately, this grid also yields some principal drawbacks. For example, there are no uniform neighborhood relations, in particular, the four horizontal and vertical neighbouring points on a grid have exactly the grid size h as distance to a considered point, while the diagonal neighbours are $\sqrt{2}h$ units away. Of course, having uniformly spaced neighbours would be a desirable property for vision purposes, since no direction would be preferred in that case. Another motivation for the utilization of different grids could be the wish for better sampling isotropy than the 90.94%, that were achieved in the rectangular or quincunx case. So as to realize these

properties, we define the *hexagonal grid*

$$\Gamma = \left\{ \left(m + \frac{1}{2} \cdot n, \frac{\sqrt{3}}{2} \cdot n \right) \mid m, n \in \mathbb{Z} \right\}.$$

A straightforward capable dilation matrix for this grid could consist of a rotation or reflection about any multiple of $\pi/3$ followed by an integer stretching about some $s \in \mathbb{N}_{\geq 2}$. This leads to dilation matrices \mathbf{Q} with $q = s^2$ or $q = -s^2/2$. However, there is a *more beautiful* one, which additionally keeps the hexagonal structure of the grid after downsampling. This *hexagonal sampling matrix* is given by

$$\mathbf{Q}_{\text{hex}} = \begin{bmatrix} 2 & -\sqrt{3} \\ \sqrt{3} & 2 \end{bmatrix}.$$

It causes a rotation of $\alpha = \arctan \frac{2}{\sqrt{3}}$ and a dilation by the factor $s = \sqrt{7}$. The sampling scheme is shown in Figure 3.5. Hereby, each seven cells within a colored hexagon are downsampled to one point in the first sampling step, while in the second step, the seven colored cells are further downsampled to one point and so on.

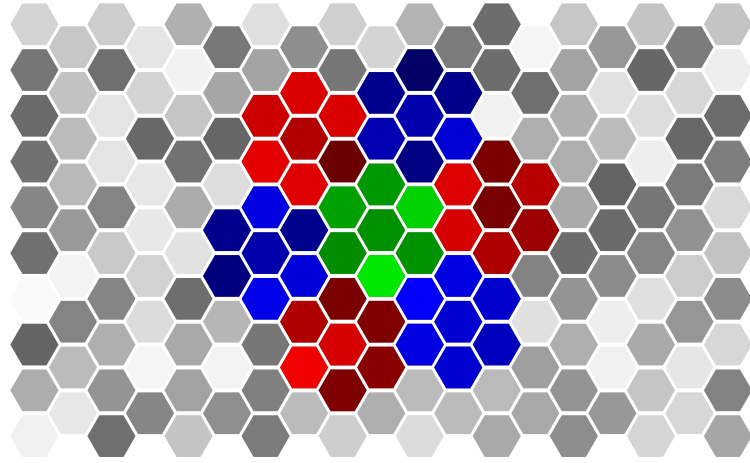


Figure 3.5. Hexagonal sampling lattice with seven channel sampling.

It is immediately clear that the VORONOI cell A_{hex} to the hexagonal sampling grid consists of an equilateral hexagon and moreover, one accomplishes 96.28% of the energy of an ideal isotropic sampling scheme again measured probabilistically via the areal coincidence ζ_{hex} between A_{hex} and a circle of the same area.

The hexagonal sampling scheme plays an important role in many biological vision processes. For example, the eyes of most mammals and insects are built out of hexagonally ordered retinal receptors. In the same context, WATSON and AHUMADA applied a hexagonal sampling scheme to create a model of image representation in the visual cortex [232] [231]. On the other hand, hexagonal sampling has also proven to be useful in non-biological but technical vision processes. In particular, LAWTON and RESNIKOFF have successfully applied nonseparable wavelets associated to this sampling scheme in a patent for optical telescopes to reduce arising GIBBS effects [135]. In [136] complex-valued HAAR wavelets for \mathbf{Q}_{hex} were built by the straightforward construction

$$\psi^m = \sum_{k=1}^7 e^{2\pi i(m-1)(k-1)/7} \cdot \varphi(\mathbf{Q}_{\text{hex}} \cdot \vec{x} - \vec{r}_k),$$

where $\{\vec{r}_1, \dots, \vec{r}_7\}$ denotes a complete set of representatives of the cosets of the lattice $\mathbf{Q}_{\text{hex}}^T \cdot \Gamma$. By solving a certain set of eight nonlinear equations of order two, we can also build real-valued wavelets for \mathbf{Q}_{hex} , these are shown in Figure 3.6. The indicator set of these functions is a fractal set known as GOSPER island.

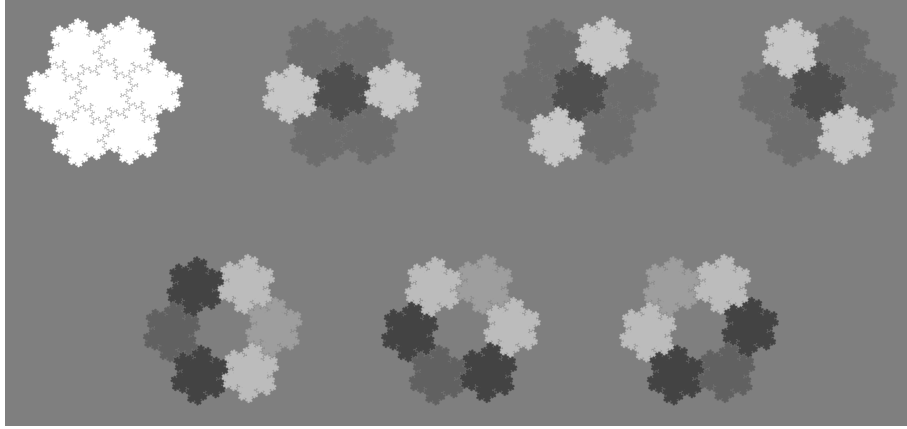


Figure 3.6. The hexagonal HAAR scaling function (upper left) and the six associated (real-valued) wavelets. Light regions indicate positive and dark regions negative function values. The three wavelets in the upper row are symmetric, those in the lower row are anti-symmetric.

Hexagonal sampling is a highly efficient scheme for image processing purposes with a big relevance for biological vision systems and moreover, it attains a very good sampling isotropy (96.28%). Unfortunately, there exist no cameras or machines working with hexagonal resolutions and data structures, except the more than 20 years old TAS[®] system by Cambridge Instruments — by personal

opinion of the author, this omission is a mistake, since the theoretical results give hope for a very effective modelling of vision processes on hexagonal grids, that are *very close* to the vision processes of the biological world and may also lead to a better understanding of the latter ones. The power of hexagonal sampling schemes was also confirmed by several experiments in [245]. Anyway, using suited linear transforms, the hexagonal grid Γ could be *rectangularized* to perform *quasihexagonal* image processing, but otherwise, this would also destroy several of its beautiful properties, for example symmetry or uniform neighborhood relations get lost this way.

We close the discussion of twodimensional fully regular sampling schemes with a brief presentation of the outstanding equilateral triangular case. Triangular sampling also *lives naturally* on the same grid Γ as above. Again, there are several obvious possibilities to implement various subsampling schemes, this shall not be discussed here. In the typical cases, the corresponding VORONOI cell consists — as desired — of an equilateral triangle. By simple calculus, one evaluates, that this scheme accomplishes only 85.22% of the ideal isotropic sampling energy and is thus much less fitted for applications than the cases discussed so far. The issue of triangular sampling lattices was discussed in more detail in [126].

3.3. Higher Dimensions

To complete the investigations about sampling in higher dimensions, we shall also discuss the general case. We start with some considerations about fully regular sampling schemes. Recall that this means to have a VORONOI cell consisting of a convex regular polytope, that gives rise for a disjoint covering (tessellation) of the considered space. These polytopes can be categorized by the following theorem borrowed from geometry:

THEOREM 3.5 (COXETER) *For all $n \in \mathbb{N} \setminus \{2, 4\}$, there exists exactly one fully regular tessellation or honeycomb of the \mathbb{R}^n , this is given by the n -cube. For $n = 2$ and $n = 4$ there exist three fully regular tessellations.*³

Proof. See [54].

□.

This result may look surprisingly at a first glance, but it isn't that much (maybe only for $n = 3$ the result is *really* surprising). Viz recalling LUDWIG

³For $n = 4$, these honeycombs are given by the 4-cube, the 4-simplex — this is the fourdimensional generalization of the equilateral triangle — and a regular polytope called *24-cell* consisting of 24 octahedrons arranged in \mathbb{R}^4 , it was discovered by L. SCHLÄFLI. More about honeycombs and regular polytopes in higher dimensions can be found in the classical book [54].

SCHLÄFLI's classical *polytope classification theorem*, for any dimension n except for $n < 5$ there exist exactly three convex regular polytopes, these are the n -cube, the n -simplex and the n -crosspolytope. Excluding the latter two cases for dimensions greater than 4, COXETER's result becomes much less a surprise. For us, the essence of this theorem is that if one is interested in fully regular sampling schemes, one has to rely on n -cubic sampling lattices as long as one does not work in dimensions two or four. Another way around could be to allow also convex *quasiregular* tessellations, i.e. using polytopes consisting of two different kind of faces but having one fixed vertex constellation (by the duality principle for polytopes, the vice versa case might also be admissible) as VORONOI cells. This alternative gives rise to some highly efficient sampling schemes that are very close to ideal isotropic sampling, in Example 2, we will present one such possibility.

3.3.1. Classification of Dilation Matrices

In order to categorize, what makes a *good* dilation matrix, we may join all the ideas presented so far in this whole chapter into a design principle for dilation matrices.

GENERAL PRINCIPLE* 3.6 *A matrix $\mathbf{Q} \in \text{Mat}(n \times n, \mathbb{Z})$ that builds the fundamental for a multidimensional MRA shall satisfy the following properties:*

- $\det \mathbf{Q} = \pm 2$ (\longrightarrow simpler wavelet construction).
- \mathbf{Q} must be strictly expansive, even better would be that all eigenvalues λ_i of \mathbf{Q} have the same modulus, i.e. $|\lambda_i| = \sqrt[n]{2}$ (\longrightarrow same dilation factor along all principal axes).
- The induced VORONOI cell shall have a great areal coincidence to an ideal circle, i.e. $\zeta_{\mathbf{Q}}$ is close to 1 (\longrightarrow nearly isotropic sampling).

Recalling the presentation so far, we know that the quincunx matrices satisfy this design principle. To illustrate the General Principle 3.6 at a higherdimensional but still vivid situation, we will give an example in 3D:

EXAMPLE* 2 (RHOMBIC DODECAHEDRAL SAMPLING) *Consider the scaling matrix*

$$\mathbf{Q}_{RD} = \begin{bmatrix} 1 & 0 & -1 \\ 1 & -1 & 1 \\ 0 & -1 & 0 \end{bmatrix}.$$

It is directly verified that \mathbf{Q}_{RD} follows the first two items of the General Principle 3.6, since

$$\det \mathbf{Q}_{RD} = 2 \quad \text{and} \quad |\lambda_i| = \sqrt[3]{2}$$

for all three eigenvalues of \mathbf{Q}_{RD} . The reader may check that the associated VORONOI cell A_{RD} is given by a rhombic dodecahedron, i.e. a dual quasiregular polyhedron consisting of 12 rhombics as faces and 14 vertices arranged in two different vertex constellations (see Figure 3.7). The areal coincidence ζ_{RD} can be evaluated to attain the very good value of 95.54% of an ideal isotropic sampling in three dimensions, which is even better than the closely related quincunx sampling in 2D. A similar sampling called FCO scheme (face centered orthorombic sampling) using a different scaling matrix was used for video coding purposes in [128].

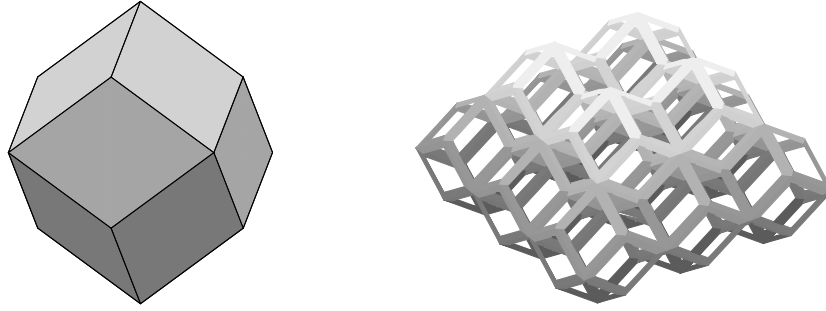


Figure 3.7. The rhombic dodecahedron and the associated tessellation of \mathbb{R}^3 .

As mentioned above, this example can be seen as the natural generalization of quincunx sampling to three dimensions. We will show that such generalized quincunx dilation matrices having two entries of modulus 1 and else zeros in every column exist for all dimensions $n \in \mathbb{N}$.

PROPOSITION* 3.7 (CLASSIFICATION OF DYADIC DILATION MATRICES)
For every $n \in \mathbb{N}$ with $n \geq 2$ there exist a dilation matrix $\mathbf{Q}^{(n)}$ that fulfills the General Design Principle 3.6 and its sampling scheme is the n -dimensional analogue to quincunx sampling. Such matrices are given by

$$\mathbf{Q}^{(n)} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & -1 \\ 1 & -1 & 1 & 1 & \cdots & 1 \\ 0 & -1 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{bmatrix}. \quad (3.14)$$

Remark. For a given n these matrices are not unique. The number v_n of different dilation matrices depends on the number of different polynomials of degree n with coefficients in $\{-2, -1, \dots, 2\}$ having all its roots on the circle

of radius 2 (note that these are not only cyclotomic polynomials). One obtains e.g. $v_2 = 3$ (see previous section), $v_3 = 4$ and $v_4 = 6$. On the other hand, it seems like there always exists a unique generalized quincunx matrix having determinant $\det \mathbf{Q}^{(n)} = -2$, but a proof for this conjecture could not be found yet.

Proof of Proposition 3.7. Since the case $n = 2$ was already explained in the previous section, we may suppose, that $n \geq 3$. As first, we can evaluate the determinant by expanding for the last column of $\mathbf{Q}^{(n)}$ and obtain

$$\begin{aligned} \det \mathbf{Q}^{(n)} &= (-1)^n \cdot \det \mathbf{Q}_{2..n,1..n-1}^{(n)} + (-1)^n \cdot \det[\mathbf{Q}_{1,1..n-1}^{(n)\top} \mathbf{Q}_{3..n,1..n-1}^{(n)\top}]^\top \\ &= (-1)^{n+1} \cdot 2 = \pm 2. \end{aligned}$$

To show that $\mathbf{Q}^{(n)}$ is expansive, we calculate its characteristic polynomial $\chi_{\mathbf{Q}}$. Denote $\mathbf{X}^{(n)} = \mathbf{Q}^{(n)} - \lambda \cdot \mathbf{Id}^{(n)}$, then $\chi_{\mathbf{Q}}(\lambda) = \det \mathbf{X}^{(n)}$. Evolving this determinant with respect to the first column, we get

$$\begin{aligned} \det \mathbf{X}^{(n)} &= (1 - \lambda) \cdot \det \mathbf{X}_{2..n,2..n}^{(n)} - \det[\mathbf{X}_{1,2..n}^{(n)\top} \mathbf{X}_{3..n,2..n}^{(n)\top}]^\top \\ &= (1 - \lambda) \cdot \det \mathbf{X}_{2..n,2..n}^{(n)} + (-1)^{n-1} \cdot \det[\mathbf{X}_{3..n,2..n}^{(n)\top} \mathbf{X}_{1,2..n}^{(n)\top}]^\top. \end{aligned}$$

The second matrix in this expression is an upper triangular matrix whose determinant is obviously equal to 1 and the determinant of the first matrix can be computed by an induction over n yielding

$$\det \mathbf{X}_{2..n,2..n}^{(n)} = (-1)^{n-1} \cdot \sum_{k=0}^{n-1} \lambda^k.$$

Inserting this into the above relation, we obtain

$$\chi_{\mathbf{Q}}(\lambda) = \lambda^n - 2.$$

Thus, all eigenvalues of $\mathbf{Q}^{(n)}$ lie on a circle of radius 2, which completes the proof. Additionally, we can conclude that $\mathbf{Q}^{(n)n} = 2 \cdot \mathbf{Id}^{(n)}$ holds true. \square .

Proposition 3.7 makes no assertions about the sampling quality in the sense of its closeness to isotropy. The reason is the following: In order to measure the isotropy, one would have to calculate the volume of the intersection of the VORONOI cell of the sampling matrix with a n -sphere of volume 2 (since this is the cells area by construction). Unfortunately, there exist even no general formulas to evaluate the volumes of higherdimensional polytopes and especially none for more complicated sets like the termed intersection sets. Therefore, we can only give the approximated numerical values for the areal coincidence ζ_n between the n -dimensional VORONOI cell and the n -sphere of the same volume for some n in the following Table 3.1.

n	2	3	4	5	6	10	20
ζ_n	90.47%	95.54%	92.9%	87%	<60%	<10%	<1%
r_n	0.7979	0.7816	0.7979	0.8240	0.8537	0.9760	1.2429

Table 3.1. Isotropy measures for generalized quincunx sampling in higher dimensions. The second row gives the radius for the n -sphere of volume 2.

3.3.2. The Isotropy Measure ζ_n for Increasing n

The worse sampling isotropy for bigger values of n may surprise a little. But it turns out that such behaviour is coercive. If the number of spatial dimensions tends to infinity, sampling becomes fully anisotropic for any dyadic dilation matrix. In particular, we can prove the following result.

THEOREM* 3.8 *Let $\mathbf{Q}^{(n)}$ be a dyadic dilation matrix for \mathbb{R}^n as in (3.14) of Proposition 3.7. Then*

$$\lim_{n \rightarrow \infty} \zeta_n = 0.$$

Proof. We will estimate the areal coincidence ζ_n from above and from below. First, it is evident that the n -volume of the VORONOI cell $A_{\mathbf{Q}^{(n)}}$ equals 2 by construction. Due to the symmetry of the dyadic sampling scheme, $A_{\mathbf{Q}^{(n)}}$ can be characterized by

$$\begin{aligned} A_{\mathbf{Q}^{(n)}} &= \left\{ \vec{x} \in \mathbb{R}^n \mid \|\vec{x}\| \leq \|\vec{x} - \vec{p}\|, \vec{p} \in \mathbf{Q}^{(n)} \cdot \mathbb{Z}^n \setminus \{0\} \right\} \\ &= \left\{ \vec{x} \in \mathbb{R}^n \mid |x_j| + |x_k| \leq 1, j \neq k \right\}. \end{aligned}$$

This representation makes it directly evident that the cube

$$C_n(\delta) = \left\{ \vec{x} \in \mathbb{R}^n \mid |x_j| \leq \delta, j = 1, 2, \dots, n \right\}$$

always lies completely within $A_{\mathbf{Q}^{(n)}}$ for $\delta \leq \frac{1}{2}$. Using some fundamental geometric calculus, one evaluates the radius of the n -sphere with volume 2 to be

$$r_n = \frac{\sqrt[n]{2 \cdot \Gamma(\frac{n}{2} + 1)}}{\sqrt{\pi}},$$

where $\Gamma(\cdot)$ denotes the EULER Gamma function. From simple calculations it then follows that the cube $C_n(\rho_n)$ with $\rho_n = r_n/\sqrt{n}$ always lies within the n -sphere $A_{n\text{-sphere}}$ of volume 2. This gives the lower estimation

$$\zeta_n \geq \frac{1}{2} \cdot \text{vol} \left(C_n \left(\min(1/2, \rho_n) \right) \right) = \frac{1}{2} \cdot \min(1, 2 \cdot \rho_n)^n.$$

Applying the factorial relation to the Gamma function and using $n! \leq (n/2)^n$ for sufficiently large numbers n , one may verify the following estimation

$$\rho_n = \frac{\sqrt[n]{2 \cdot \left(\frac{n}{2}!\right)}}{\sqrt{n \cdot \pi}} \leq \frac{\sqrt[n]{2}}{\sqrt{n \cdot \pi}} \cdot \sqrt{\frac{n}{4}} = \frac{\sqrt[n]{2}}{2 \cdot \sqrt{\pi}} \longrightarrow \frac{1}{2 \cdot \sqrt{\pi}}. \quad (3.15)$$

It can be checked that we have in fact $\rho_n < 1/2$ for $n > 2$ which means that the cube circumscribed by the n -sphere is totally intersecting with $A_{Q(n)}$. To estimate an upper value for ζ_n , we introduce the cuboid pairs

$$D_n(\delta, k) = \{ \vec{x} \in \mathbb{R}^n \mid |x_j| \leq \delta, \ j = 1, 2, \dots, n, \ j \neq k \wedge \delta \leq |x_k| \leq 1 \}.$$

Since the set $A_{Q(n)}$ is convex and $C_n(\rho_n) \cap A_{Q(n)} = C_n(\rho_n)$ for $n > 2$, we have

$$U_n = \bigcup_{k=1}^n D_n(\rho_n, k) \cup C_n(\rho_n) \supset A_{Q(n)} \cap A_{n\text{-sphere}}, \quad (3.16)$$

see also Figure 3.8 and it immediately follows the upper bound

$$\frac{1}{2} \cdot \text{vol}(U_n) \geq \frac{1}{2} \cdot \text{vol}(A_{Q(n)} \cap A_{n\text{-sphere}}) = \zeta_n. \quad (3.17)$$

To complete the proof, we have to evaluate the limit of $\text{vol}(U_n)$ as n tends

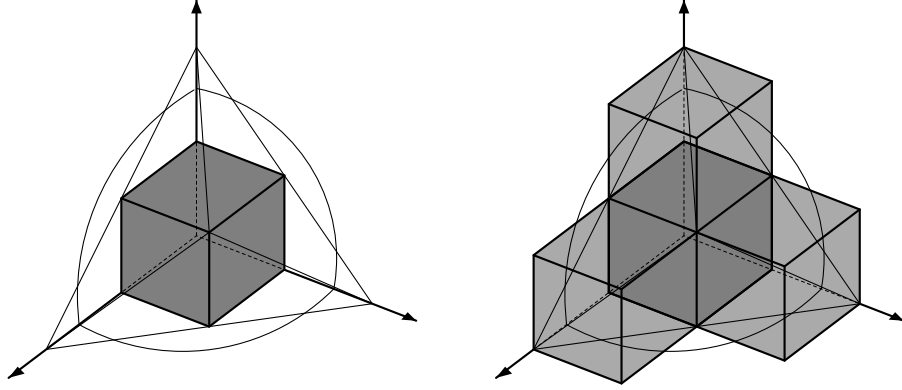


Figure 3.8. Geometrical interpretation of the sets $C_n(\rho_n)$ (left figure) and $D_n(\rho_n)$ (right figure) in (3.16) yielding an upper and lower estimation of ζ_n .

to infinity. We first insert the cuboid volumes for the sets C_n and D_n into the estimation (3.17) yielding

$$\begin{aligned} \zeta_n &\leq 2^{n-1} \cdot (\rho_n^n + n \cdot \rho_n^{n-1} \cdot (1 - \rho_n)) \\ &\leq 2^{n-1} \cdot (\rho_n^n + n \cdot \rho_n^{n-1}) \end{aligned}$$

for $n \geq 3$. Using (3.15), this gives

$$\begin{aligned}\zeta_n &\leq 2^{n-1} \cdot (2 + n \cdot 2^{2-1/n}) \cdot (2\sqrt{\pi})^{-n} \\ &= \frac{1 + n \cdot 2^{1-1/n}}{\sqrt{\pi}^n}.\end{aligned}$$

Since the numerator grows sublinear and the denominator grows exponentially, we can conclude that $\lim_{n \rightarrow \infty} \zeta_n = 0$, which is the claimed result. \square .

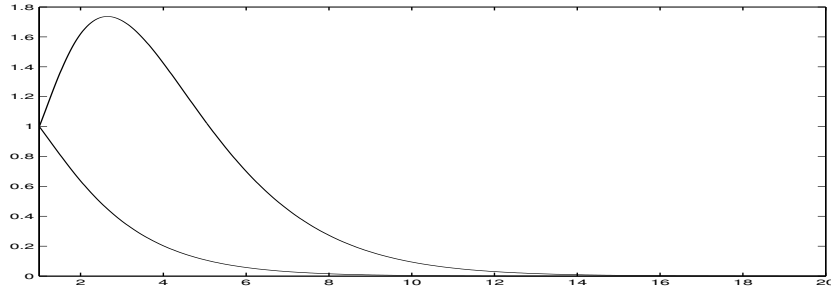


Figure 3.9. Upper and lower bounds for the areal coincidence ζ_n in the case of dyadic n -dimensional sampling.

By slightly modifying the argumentation in the previous proof, one can also obtain a more general version.

COROLLARY* 3.9 *For any dilation matrix \mathbf{Q} of an infinite-dimensional space inducing finite subsampling, i.e. $|\det \mathbf{Q}| < \infty$, the associated sampling scheme is fully anisotropic:*

$$\zeta_{\mathbf{Q}} = 0.$$

The *proof* of this result is left to the reader. We only point out that it works absolutely similar to that one of Theorem 3.8; just the helping sets C_n and D_n have to be chosen in a more general way, which makes the numerical estimations a little harder. \square .

These last two results are unpleasant from a certain point of view. But seen from a more practical position, these negative findings are not a too big obstruction, since typical wavelet applications require rather low-dimensional spaces as fundamentals. In this sense, the dyadic dilations classified in Proposition 3.7 give very good results at least for all $n < 6$. On the other hand, the reader may expect that the asymptotical behaviour of the isotropy measure ζ_n for regular sampling matrices like $\mathbf{Q} = 2 \cdot \mathbf{Id}^{(n)}$ gives better results. Unfortunately, this is also not the case.

COROLLARY* 3.10 *The n -dimensional regular sampling scheme induced by dilation matrices $\mathbf{Q} = k \cdot \mathbf{Id}^{(n)}$ becomes fully anisotropic as n tends to infinity for all $k \geq 2$.*

Proof. The principle is again the same as in the proof of Theorem 3.8. The VORONOI cell according to \mathbf{Q} is given by the n -cube of volume k^n , that is $A_{\mathbf{Q}} = C_n(k/2)$ and the radius of the n -sphere of the same volume is easily evaluated to fulfill

$$r_n = k \cdot \frac{\sqrt[n]{\Gamma(\frac{n}{2} + 1)}}{\sqrt{\pi}}.$$

Using the estimation (3.15), the statement can now be proven by the same conclusions as above. This is just some basic calculus and is thus left to the interested reader. \square .

Concluding these last findings, none of the typical sampling schemes used in dimensions 1, 2 or 3 gives rise to *nice* generalizations to arbitrary dimensional spaces. Nevertheless, as was pointed out above, for typical application spaces ($\dim \leq 6$) good dilation matrices are given by our design principle and the existence result in Proposition 3.7. We finish this section with the remark, that it is not clear at all whether families of dilation matrices exist such that their scaling isotropy ζ_n does *not* tend to zero as n goes to infinity.

3.4. Extensions and Related Work

The notion of a multiresolution analysis can be generalized in many different ways. The direct n -dimensional extension presented in Section 1 is used here since this setting is sufficiently suited for the applications we are targeting at. A very elaborate work with detailed explanation and many illuminating examples of such multidimensional MRAs is [143]. Constructions of this kind were further generalized to *multidimensional multiwavelet MRAs* in [32] [34] [31], leading to subspace sequences formed from multiple scaling functions. This approach allows even symmetric orthogonal wavelets which were inaccessible in the case of *standard* MRAs. Moreover, the wavelets could be designed to be piecewise polynomial in 1D. Onedimensional multiwavelets were successfully applied to numerical solutions of advection-diffusion equations [2].

Another much more general extension of MRAs is that coming along with the so-called *lifting scheme* [216] [218] of W. SWELDENS. This can be applied completely detached from any fixed dilation matrix and allows one to build wavelets on arbitrary complete metric spaces. Hereby, the dilation is (mostly) maintained by triangulations of the considered space and newly chosen filters on each scale. Lifting allows simple construction of biorthogonal filters, but full orthogonality is not possible in general. Due to its flexibility, lifting is

especially well-suited for numerical applications on non-regular domains, but the huge technical machinery coming along with domain decompositions and establishing of regularity [66] [65] [100] makes it less attractive for image processing, where one mainly deals with simple rectangular (or cuboidal) data sets. A similar formulation to the lifting scheme can also be found in the work of DAHMEN and his co-workers working on wavelet-based numerical solutions for PDEs [40] [58].

Much less work has been done so far regarding the design or classification of scaling matrices. Most considerations were focussing on the regular or the quincunx sampling in 2D, the *close-to-isotropic* hexagonal sampling in 2D or any higherdimensional cases were investigated quite rarely. The author does only know about one work, where the general case was directly addressed and this was in a very rudimentary manner. Namely, in [127], two classes of scaling matrices for arbitrary dimensions were introduced, but while one of these classes was not even expansive, the other class always induced preferred scaling directions, which is in general not desired. Our Proposition 3.7 and Theorem 3.8 give the first sensible classification of dyadic dilation matrices in arbitrary dimensions together with an evaluation of the sampling quality that comes along with these matrices. Additionally, as we argued above, we believe that dilations for higher dimensions should strictly follow the General Design Principle 3.6.

Chapter Summary

The multiresolution analysis concept, the basic relations defining discrete wavelets in higher dimensions as well as the meaning of dilation matrices were discussed at the beginning of this chapter. In the sequel, some typical twodimensional sampling schemes were investigated and compared to each other. Afterwards, we have considered the general multidimensional case and introduced some principal guidelines how a dilation matrix for an arbitrary dimensional space should look like. We have also shown the existence of an infinite family of multidimensional dyadic scaling matrices which generalize the important quincunx sampling scheme. The chapter was closed by a more theoretical investigation about the asymptotic behaviour of the sampling isotropy for large classes of dilation matrices with increasing dimensions.

Chapter 4

DESIGNING WAVELETS

Design is an art.

—GIORGIO ARMANI

In the multidimensional case, there are two main factors that influence the properties of nonseparable wavelets. The first is the dilation matrix used to build up a multiresolution analysis; this issue was thoroughly discussed in the previous chapter. The second, even more important thing is the discrete filter that implies the spatial and spectral properties of the resulting wavelets. Consequently, we will now present criteria and methods for multidimensional wavelet filter design.

The actual chapter is organized as follows. In the first section, a formal generalization of some important mathematical relations from onedimensional wavelet theory is given as well as a new type of characterization of multidimensional filters in terms of lowerdimensional *subfilters* which are implicitly contained in these. Furthermore, we will sketch how the concept of basic filters is generalizable to higher dimensions. Once, this is done we argue that the more complicated nonseparable wavelets have in fact some principal advantages over their tensor product counterparts and are thus indeed considerable for applications. The second chapter deals with several different types of orthogonal wavelets. Starting with the presentation of general design principles, different examples of multidimensional wavelets representing the implementation of different features are presented. Afterwards, we discuss the advantages and differences to other filter design approaches. The short third section gives an overview over the concept of biorthogonal wavelets in more than one dimension. It follows the presentation of the idea of tight wavelet frames, which

make it possible to combine the features of orthogonal wavelets (stable representations, inversion by adjoint operators) with other features as symmetry and higher smoothness which are somewhat contrary to the concept of orthogonal wavelets. This section also implies a construction of arbitrary regular tight wavelet frames in any dimension obtained by a rather simple combinatorial lemma. Finally, the chapter is closed by a discussion about which kind of wavelets shall be applied in what kind of application.

4.1. Conceptual Generalization

To build wavelets in arbitrary dimensions, we do not only need a generalized concept for the necessary sampling and filtering operations, but also a complete new way for the design procedure for scaling and wavelet filters is required. This goes back to the fact that some of the convenient properties of the onedimensional case like the factorability of the transfer function are no longer valid, see also the results mentioned in Paragraph 2.2.2. Consequently, this section shall build the fundamental to implement the suited methods that are necessary to design multidimensional wavelets and to extend the direct design method also sketched in Paragraph 2.2.2. Afterwards, we will argue, why we propose to use nonseparable multidimensional wavelets instead of the simpler tensor product wavelets. This discussion will be supported by some simple examples from image processing.

4.1.1. Some Fundamental Relations

We start the development of strategies for multidimensional wavelet filter design with the prerequisite, that we are interested in orthogonal wavelets. This restriction at the early stage is useful to explain the mechanisms in a clear way, since the orthogonal case is mathematically paradigmatic for this problem — nevertheless, in the Sections 4.3 and 4.4 we will introduce generalizations to non-orthogonal wavelets, which will also be profitable in certain situations. So, the first thing to be done is to generalize the orthogonality constraint (2.14) to higher dimensions. This is done in the following matter.

LEMMA 4.1 *Let $\{\vec{0} = \vec{r}_1, \vec{r}_2, \dots, \vec{r}_{|q|}\}$ be a complete set of representatives of the cosets of the reciprocal sampling lattice $\Gamma = \mathbf{Q}^T \cdot \mathbb{Z}^n$ in \mathbb{Z}^n , that is*

$$\mathbb{Z}^n = \bigcup_{j=1}^{|q|} (\vec{r}_j + \Gamma),$$

where $\mathbf{Q} = (q_{i,j})$ again denotes a scaling matrix with $\det \mathbf{Q} = q$. Suppose now that $\varphi(\vec{x})$ satisfies a scaling equation (3.7) with filter coefficients $h_{\vec{k}}$, $\vec{k} \in \mathbb{Z}^n$ such that $\{\varphi(\cdot - \vec{k})\}_{\vec{k} \in \mathbb{Z}^n}$ forms an orthonormal system. Then the transfer

function given by (3.9) must fulfill the orthogonality constraint

$$\sum_{j=1}^{|q|} |H(\vec{\omega} + 2 \cdot \pi \cdot \mathbf{Q}^{-T} \cdot \vec{r}_j)|^2 \equiv 1 \quad \text{almost everywhere.} \quad (4.1)$$

As in Proposition 2.7, this relation is equivalent to

$$\sum_{\vec{k} \in \mathbb{Z}^n} h_{\vec{k}} \cdot h_{\vec{k} + \mathbf{Q} \cdot \vec{m}} = \delta_{\vec{m}}. \quad (4.2)$$

Proof. See [184], Chapter 7. \square .

Filters with a dyadic dilation matrix, whose transfer functions satisfy the condition (4.1) are also called *quadrature mirror filters* (QMF) in the engineering literature. Again, we remind the reader, that (4.1) and (4.2) are only necessary conditions and an additional criterion like the COHEN or LAWTON criterion has to be satisfied to guarantee L_2 -existence and orthogonality. Furthermore, from the requirement $H(\vec{0}_n) = 1$, we can directly deduce that the useful relations

$$H(2 \cdot \pi \cdot \mathbf{Q}^{-T} \cdot \vec{r}_j) = \delta_{1,j} \quad (4.3)$$

must hold true for all representatives \vec{r}_j . These equations may — as in the onedimensional case — be interpreted as zeroth order vanishing moments for the associated wavelet. Even more interesting is the fact that we can conclude that n -dimensional orthogonal filters may contain *subfilters* of lower dimensions $m < n$ which are also orthogonal:

LEMMA* 4.2 Suppose $H(\vec{\omega})$ with $\vec{\omega} \in \mathbb{R}^n$ is the transfer function of an orthogonal filter with a generalized quincunx matrix \mathbf{Q} . Let $\vec{\sigma} : \mathbb{R} \rightarrow \mathbb{R}^n$ denote any n -fold sign permutation, that is either $\sigma_i(x) = x$ or $\sigma_i(x) = -x$ for all $i = 1, \dots, n$. Then, the symbol $\tilde{H}(\vec{\omega})$, $\vec{\omega} \in \mathbb{R}$ with

$$\tilde{H}(\vec{\omega}) = H(\vec{\sigma}(\vec{\omega})) \quad (4.4)$$

is the transfer function of an onedimensional orthogonal filter with integer dilation factor 2.

Proof. We have $\tilde{H}(\vec{\omega}) = H(\sigma_1(\vec{\omega}), \dots, \sigma_n(\vec{\omega}))$ and due to the periodicity of the complex exponential function, we also have

$$e^{i \cdot \sigma_j(w + \pi)} = e^{i \cdot (\sigma_j(w) + \pi)}$$

for all $j \in 1, \dots, n$. Recalling Proposition 3.7, the set $\{\vec{0}, \vec{e}_1\}$ is obviously a complete set of representatives of the cosets of the reciprocal lattice $\Gamma = \mathbf{Q}^T \cdot \mathbb{Z}^n$.

Together with Lemma 4.1 and the fact that $H(\vec{\omega})$ is a trigonometric polynomial this gives us

$$\begin{aligned}
 |\tilde{H}(\vec{\omega})|^2 + |\tilde{H}(\vec{\omega} + \pi)|^2 &= |H(\vec{\sigma}(\vec{\omega}))|^2 + |H(\vec{\sigma}(\vec{\omega} + \pi))|^2 \\
 &= |H(\vec{\sigma}(\vec{\omega}))|^2 + |H(\vec{\sigma}(\vec{\omega}) + \vec{\pi})|^2 \\
 &= |H(\vec{\sigma}(\vec{\omega}))|^2 + |H(\vec{\sigma}(\vec{\omega}) + 2 \cdot \pi \cdot \mathbf{Q}^{-T} \cdot \vec{e}_1)|^2 \\
 &= 1 \quad \text{almost everywhere.}
 \end{aligned}$$

Since $H(\vec{\omega})$ leads to an orthonormal scaling function, it must satisfy COHENS criterion. By construction, the same must be true for $\tilde{H}(\vec{\omega})$ and this guarantees the existence of an orthonormal scaling function in 1D. \square .

In the same manner, one may show that in the generalized quincunx case, there exist orthogonal subfilters of any dimension $m < n$ by replacing the sign permutation $\vec{\sigma}$ by a suited sign-and-variable permutation

$$\begin{aligned}
 \vec{\tau} : \mathbb{R}^m &\longrightarrow \mathbb{R}^n \\
 (x_1, \dots, x_m) &\longmapsto \begin{pmatrix} \pm x_{i_{11}} \pm x_{i_{12}} \dots \pm x_{i_{1k_1}} \\ \vdots \\ \pm x_{i_{n1}} \pm x_{i_{n2}} \dots \pm x_{i_{nk_n}} \end{pmatrix}.
 \end{aligned}$$

Subfilters do not only exist in the dyadic case. Under some additional conditions, their existence may also be shown for several other dilations. In the following, we will show such a result for a regular sampling matrix.

LEMMA* 4.3 *Suppose $H(\vec{\omega})$ with $\vec{\omega} \in \mathbb{R}^n$ is the transfer function of an orthogonal filter for the regular scaling matrix $\mathbf{Q} = [2 \cdot \delta_{j,k}]_{j,k}$. Then, any onedimensional symbol function $\tilde{H}(\omega)$ with*

$$\tilde{H}(\vec{\omega}) = H(0, \dots, 0, \vec{\omega}, 0, \dots, 0) \quad (4.5)$$

is the transfer function of an onedimensional orthogonal filter with dilation factor 2 if and only if $H(\vec{\omega}) = 0$ for any $\vec{\omega} \in \mathbb{R}^n$ such that $\omega_j = \pi$ for at least one index $j \in 1, \dots, n$.

Remark. The additional condition may look very restrictive at a first glance. But since the scaling filter is usually interpreted as a lowpass filter, it is quite natural to demand the symbol to vanish not only at the aliasing frequencies but also anywhere on the highpass boundaries which are just given by those frequencies $\vec{\omega}$ which have at least one component ω_j attaining the value π .

Proof. Since we have $\det \mathbf{Q} = 2^n$ and a regular sampling lattice, the cosets of the reciprocal lattice are given by all elements of $S_n(\{0, 1\})$ where $S_n(\Lambda)$

denotes the set of all possible vector combinations of n not necessarily different elements of Λ . Assuming the additional condition, we are allowed to add the $2^n - 2$ zero terms in the following calculation.

$$\begin{aligned}
|\tilde{H}(\tilde{\omega})|^2 &+ |\tilde{H}(\tilde{\omega} + \pi)|^2 \\
&= |H(0, \dots, 0, \tilde{\omega}, 0, \dots, 0)|^2 + |H(0, \dots, 0, \tilde{\omega} + \pi, 0, \dots, 0)|^2 \\
&= \sum_{\vec{r} \in S_n(\{0,1\})} |H((0, \dots, 0, \tilde{\omega}, 0, \dots, 0) + \pi \cdot \vec{r})|^2 \\
&= \sum_{\vec{r} \in S_n(\{0,1\})} |H((0, \dots, 0, \tilde{\omega}, 0, \dots, 0) + 2 \cdot \pi \cdot \mathbf{Q}^{-T} \cdot \vec{r})|^2 \\
&= 1 \quad \text{almost everywhere.}
\end{aligned}$$

The proof can again be concluded by the remark that the validity of the COHEN criterion is preserved by our construction. \square .

Lemma 4.2 and Lemma 4.3 are quite useful for the design of higherdimensional orthogonal filters, since they tell us that any filter contains simpler filters of lower dimension that might be easier obtainable. On the other hand, these lemmata do *not* give a constructive method to build higherdimensional filters directly from lowdimensional ones — this seems to be impossible due to the nonseparability, which prevents direct filter decompositions. But the knowledge can and will help us in building multidimensional filters by providing information about how the filters must look like and by reducing the algebraic complexity of the design procedure itself. All this will become clearer within the discussion in the following sections. We additionally mention that recently, a very special case of Lemma 4.3 was formulated in [105] to show that wavelets with regular dilation matrices in higher dimensions can never exceed the regularity level of their (bi)orthogonal onedimensional counterparts with maximally flat scaling filters.

4.1.2. Basic Filters in Higher Dimensions

Two questions naturally arise when one wants to generalize the concept of decomposing filters into basic filters that was presented at the end of Chapter 2. The first concerns the *optimal shape* of the filters to be designed. Hereby, optimality shall mean to find the best possible pay-off between filter shape and size (number and arrangement of taps) and number of orthonormality constraints as given by (4.1) or (4.2). The second is the question of *how* to build the basic filters in a (preferably) simple way. Both questions shall be discussed in this paragraph.

First, we will care about the filter shape. As we learned from Lemmata 4.2 and 4.3, orthogonal multidimensional filters contain orthogonal subfilters of

lower dimension. These subfilters can be obtained by summing the filter taps along certain *discrete hyperplanes* as one sees from the defining relations (4.4) and (4.5). Since onedimensional filters obviously have a unique shape, namely a line arrangement, the onedimensional subfilters of any multidimensional filter must be given by projections of the hyperplanes onto lines. This means that the optimal filter shape is that, for which the projections of the hyperplanes of the multidimensional filter give lines of minimal length. This fact shall be illustrated in the following:

EXAMPLE* 3 (OPTIMAL SHAPES FOR QUINCUNX FILTERS) *We choose a dummy filter $[h_{j,k}]$ and $\mathbf{Q} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$ to be a quincunx dilation matrix. Following the results of this chapter, any onedimensional subfilter is given by a summation of the coefficients $h_{j,k}$ along the two diagonals $j = \pm k$. This means that the filter shall be rhombic or diamond shaped. Since we also have a dyadic sampling, the projection of the diagonals onto lines must yield lines of even length and therefore, the filter taps of a quincunx filter are ideally arranged in a number of double diagonals each of the same length. Consequently, taps of this shape can be found anywhere (but generally without any detailed justification) in the literature about wavelets based on quincunx matrix dilations [129] [169] [212].*

The answer to the second question is pretty simple. Knowing the optimal shape of the basic filters, one has to build a collection of fundamental filters whose repeated convolution yields exactly this shape and such that the collection contains a filter of each of the 2^n possible different symmetries in n -dimensional space, since every (transfer) function can be uniquely decomposed into its symmetric and anti-symmetric parts with respect to the axes; this guarantees the completeness of the filter family. Additionally, the fundamental filters must satisfy *sum rule* conditions like (2.18) to ensure that repeated convolutions of the fundamental filters lead to higher order basic filters to get a similar situation to that of Lemma 2.13 and Corollary 2.14. Actually, this is all, just a little flair is needed to find these fundamental filters. Building basic filter families will be demonstrated in more detail in Subsection 4.2.2.

4.1.3. Why Using Nonseparable Multidimensional Wavelets?

So far we have seen that the design of nonseparable wavelets is significantly harder than building tensor product wavelets from their well-known onedimensional prototypes and therefore, one should ask for the sensibility of using nonseparable wavelets at all. In other words, we must ask the question, whether the harder work of building nonseparable wavelets and working with them is really worth while. Consequently, it will be argued now that these wavelets have in fact some principal advantages in comparison to tensor product wavelets and

are thus attractive for applications.

By construction, tensor product wavelet transforms yield preferred directions along the spatial axes, which leads to inhomogeneous sensitivity against structures within the data and thus gives undesired processing artefacts in most applications. In contrast to this, nonseparable wavelets give a *richer structural description* in the sense that they are principally closer to homogeneity, because they have no such preferred processing directions. This fact is also demonstrated by the example shown in Picture 4.1.

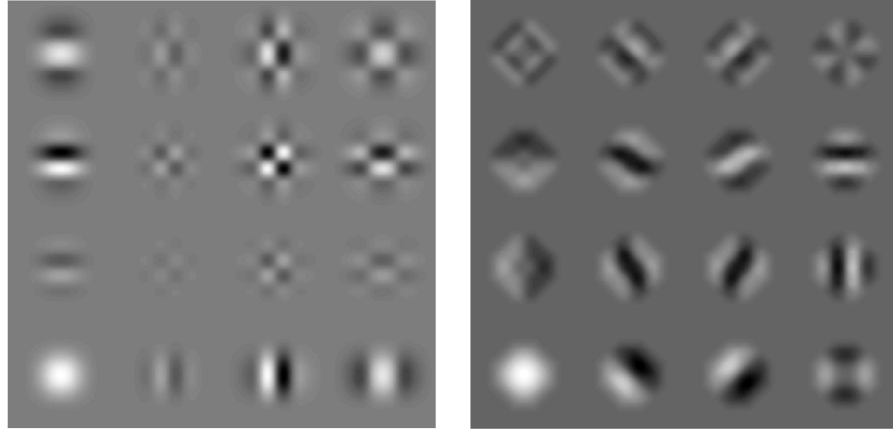


Figure 4.1. Spatial resolution of wavelet frames with two vanishing moments. *Left.* Separable case. *Right.* Nonseparable case. While the tensor product filters are only sensitive with respect to the axes, the nonseparable counterparts are also receptive against rotations.

Let us now consider the total number of free variables in the case that the filters are separable and nonseparable. In the first case, a n -dimensional filter has

$$\#V_s = \sum_{i=1}^n (k_i - 1)$$

degrees of freedom where the k_i denote the filter length in the i -th dimension. Note that the subtraction of 1 is for normalization reasons. Now, if the filter is nonseparable, the number of free variables is

$$\#V_n = \prod_{i=1}^n k_i - 1,$$

which makes it evident that

$$\#V_n > \#V_s. \quad (4.6)$$

This means that one has (for bigger filters significantly) *more design freedom*. These degrees of freedom can be utilized to implement several desired features into nonseparable filters, which are inaccessible with separable filters.

One example of such a desired feature is to build low-pass filters such that the associated scaling functions are close to radial symmetric functions. This feature is somehow connected to the abovementioned richer structural description, since it yields *close-to-isotropic* data sampling and thus improves the homogeneity of the wavelet transform. Another important feature that could be implemented using the additional degrees of freedom is to optimize the energy compaction of the low-pass filter in frequency domain in order to compress data representations or to reduce the computational amount of processing tasks. The concrete realization of such a feature will be presented in the paragraph about filter optimization within the next section.

There are even more reasons to prefer nonseparable wavelets rather than tensor product ones. It is a commonly known fact, that wavelets have a very close connection to operators on function spaces (see [157] [159]). This connection will play an important role in the considerations of Chapter 6. At this stage, we just point out exemplarily that differential operators, that stem from tensor product wavelets correspond to intrinsically onedimensional finite difference schemes, while their nonseparable counterparts in fact correspond to multidimensional schemes and are thus better capable to register the intrinsically multidimensional local behaviour of the signal under consideration.

All this justifies the employment of nonseparable wavelets for applications. A discussion about the differences between separable and nonseparable filters that mainly takes the aspect of polyphase representations into account can be found in the fundamental article [127].

4.2. Orthogonal Wavelets in \mathbb{R}^n

This must have been the schedule's only occasion for drifting into reverie – there would seem to have been no other room for speculations, dreams, fantasies, fiction. Life in that orthogonal machine was supposed to be nonfiction.

—THOMAS PYNCHON

This section is devoted to the design of nonseparable orthonormal wavelets in arbitrary dimensions. Since the fundamental relations and design mechanisms were already presented, we mainly discuss the general criteria that should be taken into account when building scaling (and wavelet) filters. The presentation will be supported by several examples involving different features. Some of

these features will be discussed in greater detail in the paragraph dealing with optimization strategies for scaling filters. The section is closed by a brief discussion and comparison to other approaches for orthogonal filter design in higher dimensions.

4.2.1. General Design Principles

What makes a good filter? This question must stand at the very beginning of any filter design procedure and there are in fact many different thinkable properties one might like to integrate. Some of these need extra requirements, some are (at least partially) excluding others and it is definitely impossible to unite all *nice* features into one filter. But anyway, it is useful to set up some general guidelines one should always follow. First, there is one criterion that should be fundamental for any filter independently of the purpose one is targeting at, namely to have a certain number of vanishing moments for the associated wavelets. These vanishing moments make sure that all polynomials up to the same degree are already contained in the base space V_0 of the MRA. This *compactifies* representations of functions in wavelet bases and therefore reduces the amount for storage and/or computation in applications. We define

DEFINITION 4.4 *Let $f \in L_2(\mathbb{R}^n)$ be any function and $\vec{m} \in \mathbb{N}_0^n$. Then*

$$\mu_{\vec{m}, \vec{k}}^f = \int_{\mathbb{R}^n} f(\vec{x} - \vec{k}) \cdot \vec{x}^{\vec{m}} d\vec{x}$$

is called the \vec{m} -th moment of f shifted by the moment shift \vec{k} . When no confusion can occur, we use the simplification

$$\mu_{\vec{m}}^f = \mu_{\vec{m}, \vec{0}}^f.$$

Recalling Lemma 2.9 and the relation (4.3), we can give the following characterization of the vanishing moment property:

PROPOSITION 4.5 *Use the notation of Lemma 4.1 and let $\{\vec{0} = \vec{t}_1, \dots, \vec{t}_{|q|}\}$ be a complete set of representatives of the cosets of $\Lambda = \mathbf{Q} \cdot \mathbb{Z}^n$ in \mathbb{Z}^n . Then the following statements are equivalent:*

- 1 *All associated wavelets $\psi^j(\vec{x})$, $j = 1, \dots, |q| - 1$ have at least ν vanishing moments, that is*

$$\mu_{\vec{m}}^{\psi^j} = 0, \quad \forall |\vec{m}| \leq \nu. \quad (4.7)$$

- 2 *For all $j = 2, \dots, |q|$ and $\vec{m} \in \mathbb{N}^n$ with $\sum_{k=1}^n m_k \leq \nu$ the generalized STRANG-FIX conditions*

$$\frac{\partial^{\sum m_k}}{\partial \vec{\omega}^{\vec{m}}} H(2 \cdot \pi \cdot \mathbf{Q}^{-T} \cdot \vec{r}_j) = 0 \quad (4.8)$$

hold true.

3 For all $\vec{m} \in \mathbb{N}^n$ with $\sum_{k=1}^n m_k \leq \nu$ the multidimensional sum rules

$$\sum_{\vec{k} \in (\vec{t}_1 + \Lambda)} \vec{k}^{\vec{m}} \cdot h_{\vec{k}} = \dots = \sum_{\vec{k} \in (\vec{t}_{|q|} + \Lambda)} \vec{k}^{\vec{m}} \cdot h_{\vec{k}} = |\det \mathbf{Q}|^{-1} \quad (4.9)$$

are valid.

Proof. The equivalence between (4.8) and (4.9) can be easily deduced by inserting (3.9) into (4.8) and using (4.3) and Lemma 4.1. Proving (4.7) \iff (4.9) is nontrivial and very space consuming. A proof can be found in [34]. \square .

The required number of vanishing moments varies between different applications, but it is a meaningful entity anyway. For example, in image compression wavelets with two vanishing moments seem to be the optimal compromise between polynomial reproducibility on one hand and filter size (decorrelation and ringing effects) on the other [8]. In applications in numerical analysis a higher number like five or six vanishing moments can yield better results for e.g. sparse operator representations [21], while multiscale preconditioning again seems to work optimal with only two or three vanishing moments [131]. Finding the optimal value for a certain application is mostly a matter of experience and some trial-and-error.

The first (but not the only) fundamental thing wavelets are used for in image processing is *representation*. In order to obtain visually good image representations and accessible norm estimations (see Appendix B, Proposition B.4), the utilized wavelets must have some degree of smoothness or regularity. Researchers working on image coding agree that a regularity order between 1 and 2 (depending on the type of image) in terms of the HÖLDER exponent is the optimal value to achieve visually satisfying and efficient representations for images.¹ This can be motivated by the fact that the best performing wavelets for image coding all have regularity values in this area; for example, the famous (and up till today most successful) 9/7 tap of [8] gives a smoothness value of 1.701 for the synthesis functions. Another justification is the intrinsic smoothness of an image in terms of its BESOV norm [75], that seem to give perception quality measures which agree well with the capabilities of the human visual system [41]. It was shown by several experiments in [38] that typical natural images have a maximum local regularity of about 1.6, which confirms the claim.

¹Smoothness orders greater than 2 seem to have no remarkable improving effect on image codes [226], since higher order smoothness requires rapidly increasing filter lengths. But long filters yield ringing and decorrelation effects which are obviously undesirable in image coding, since they have to be compensated by greater codelengths.

Nevertheless, we should remind the reader to read this *argumentation* with care, because we did something like *comparing apples and oranges* by switching between HÖLDER and BESOV regularity; the plausibility for this can be found in the Appendix about function spaces. Coming back to the issue of filter design now, we need a technique to estimate the smoothness of the wavelets for a given filter in order to measure its suitability for image processing purposes. This is maintained by the following theorem that generalizes the results in [69] [47].

THEOREM 4.6 *Again using the notation of Lemma 4.1, we define the transition operators*

$$(\mathbf{T}_i)_{\vec{j}, \vec{k}} = h_{\mathbf{Q} \cdot \vec{j} - \vec{k} + \vec{t}_i}, \quad \vec{t}_i \in \mathbf{Q} \cdot \mathbb{Z}^n$$

with $i = 1, \dots, |q|$. These operators act on a discrete set $F \subset \mathbb{Z}^{2n}$ given by those vectors \vec{j}, \vec{k} such that all the operators \mathbf{T}_i are non-singular. Suppose now, that conditions (4.9) are valid for a certain $\nu \in \mathbb{N}$. Then for each $j \leq \nu$, $k \leq |q|$ there exist $\binom{n+j-1}{j}$ eigenvalues of \mathbf{T}_k^T with $|\lambda_{i,j}| = |q|^{-j/n}$ for all $i = 1, \dots, \binom{n+j-1}{j}$ and define $E_{i,j}$ to be the eigenspace of $\lambda_{i,j}$. If E denotes the orthogonal complement of all $E_{i,j}$, $j \leq \nu$ in the discrete set F , that is

$$F = E \oplus \bigcup_{j \leq \nu, i \leq \binom{n+j-1}{j}} E_{i,j}$$

and there exist $\lambda < 1$ and $C > 0$ such that

$$\left\| \prod_{k=1}^m \mathbf{T}_{\epsilon_k} \Big|_E \right\| \leq C \cdot \lambda^m \quad (4.10)$$

for all possible combinations of $\epsilon_k \in \{1, \dots, |q|\}$ and all $m \in \mathbb{N}$, then the solution $\varphi(\vec{x})$ of the scaling equation (3.7) is ν times continuously differentiable and all its ν th derivatives are HÖLDER continuous with critical exponent

$$\alpha = -\frac{\log \lambda}{\log |q|^{1/n}}.$$

Proof. The complete proof of Theorem 4.6 would be very lengthy and technical. Since it generalizes results in [69] and [47], we refer to these articles, where this more general case is at least outlined. The main idea of the proof is to show that points $\vec{x}, \vec{y} \in \mathbb{R}^n$ which are *close* have (at least one among a number of possibilities) similar \mathbf{Q} -adic representations. From these representations one may evaluate their function values $\varphi(\vec{x}), \varphi(\vec{y})$, which are then given by infinite products of the transition operators \mathbf{T}_i (depending on the \mathbf{Q} -adic representations of \vec{x} and \vec{y}). The decisive point is now that these infinite products coincide in the first m factors due to their similar \mathbf{Q} -adic representation. Since

the smoothness for φ is thus given by the spectral properties of the collection of transition operators, one obtains an estimation like (4.10) and this yields the desired HÖLDER regularity characterization. \square .

Remark. The question of determining the regularity of wavelets or, more generally, any refinable functions, has been a very active area of mathematical research within the last decade. The first results for onedimensional refinable functions were obtained by DAUBECHIES [60] and EIROLA [81] and were subsequently improved in the work of VILLEMOES [227], RIOUL [191] and DAUBECHIES and LAGARIAS [69], where the first special results for the multidimensional case were outlined. A special case of Theorem 4.6 was shown in [47] for checking the continuity in the quincunx case. A very similar result to that one can also be found in [228]. There exist many more approaches to estimate the smoothness of refinable functions in terms of different regularity statements like the membership in some SOBOLEV or LIPSCHITZ space [48] [117] [196]. These latter works yield somewhat weaker results than those determining the critical HÖLDER exponent, but on the other hand, these are usually easier to implement and also allow some asymptotic characterizations of smoothness, which are inaccessible with methods like Theorem 4.6.

Theorem 4.6 states that the HÖLDER smoothness of any refinable function can be estimated by the joint spectral radius of the transition operators $(\mathbf{T}_i)_{\vec{j}, \vec{k}}$. Since an exact computation of the joint spectral radius would require an infinite number of matrix multiplications, one has to rely on estimations via upper bounds.² A technique to obtain such upper bounds for the joint spectral radius of a collection of operators was developed in [95] and we used this technique to determine the regularity of all those wavelets we will present in the following paragraphs.

4.2.2. A First Example

Now, the time has come to give a first simple example of wavelet filter design to illustrate the developments presented so far. Since this is the best-suited non-standard case, we will again consider the quincunx sampling grid. First, we have to find suited basic filters to start with the design procedure. As we have already explained in Example 3, the ideal shape for a quincunx scaling filter is that of a rhombic consisting of m_1 double diagonals each of length m_2 and therefore, suited basic filters representing all possible symmetries in two dimensions may be built from the fundamental filters

²The lower bound is obviously given by the largest eigenvalue of all operators $(\mathbf{T}_i)_{\vec{j}, \vec{k}}$.

$$h_1 = \begin{bmatrix} & 1 & \\ 1 & 4 & 1 \\ & 1 & \end{bmatrix}, \quad h_2 = \begin{bmatrix} & -1 & \\ 1 & 0 & 1 \\ & -1 & \end{bmatrix}, \quad h_3 = \begin{bmatrix} & -1 & \\ 1 & 0 & -1 \\ & 1 & \end{bmatrix}$$

$$\text{and} \quad h_4 = \begin{bmatrix} & -1 & \\ -1 & 0 & 1 \\ & 1 & \end{bmatrix}.$$

While h_1 and h_2 satisfy the multidimensional sum rules of first order, the antisymmetric filters h_3 and h_4 fulfill the sum rules of only zeroth order. Because orthogonal filters for the quincunx grid have to be of even length, we should additionally maintain the HAAR filters

$$g_1 = \begin{bmatrix} 1 & 1 \end{bmatrix} \quad \text{and} \quad g_2 = \begin{bmatrix} 1 & -1 \end{bmatrix}$$

in order to get filters of the *correct* shape. Suppose now, that one wants to design a twodimensional orthogonal filter for the quincunx grid such that this additionally leads to a wavelet with one vanishing moment. Then, all convolved filters $h_i * g_j$ satisfying at least the sum rules of first order, build the family of basic filters for the design procedure. In particular, these are

$$\{ h_1 * g_1, h_1 * g_2, h_2 * g_1, h_2 * g_2, h_3 * g_1, h_4 * g_1 \}.$$

Building a linear combination of these basic filters and solving for the orthogonality constraint (4.2) under consideration of Lemma 4.2, one obtains the solutions

$$h = \frac{1}{16} \cdot h_1 * g_1 + \frac{1}{8} \cdot h_2 * g_1 \pm \frac{\sqrt{3}}{16} \cdot h_1 * g_2 \pm \frac{\sqrt{3}}{8} \cdot h_4 * g_1, \quad (4.11)$$

which reproduces the famous KOVAČEVIĆ-VETTERLI scaling filter (see also [127]), the first known orthogonal filter that leads to a continuous wavelet for the quincunx dilation matrix $\mathbf{Q}_{\text{quin},2}$. Higher order filters can be built in a similar manner using basic filters consisting of longer convolutions in order to achieve more vanishing moments.

Finally, we shall illustrate Lemma 4.2 by means of this first example by considering one particular solution of (4.11). In Figure 4.2 one directly sees, how the onedimensional subfilters (which must be and in fact are the DAUBECHIES tap-4 filter in this case) are contained in the original filter in the quincunx case.

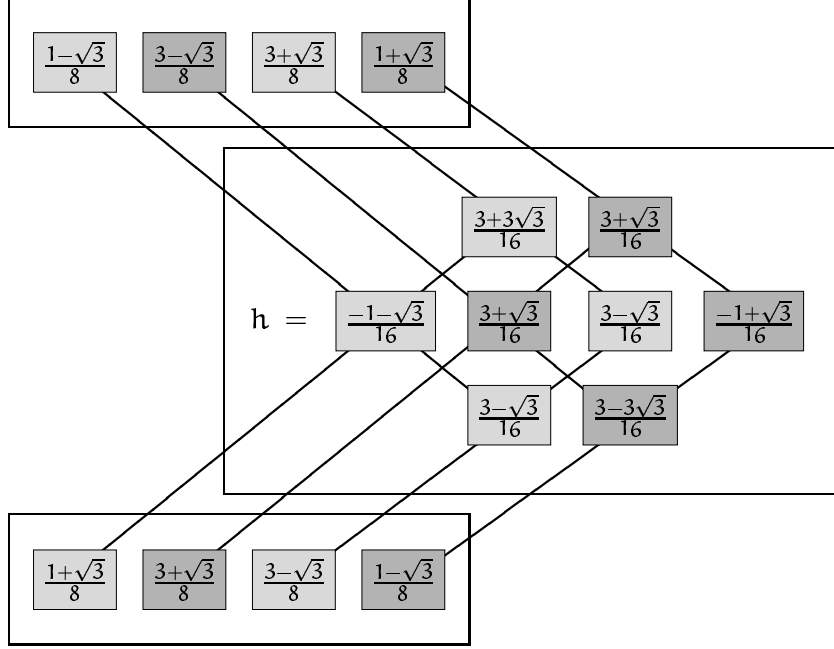


Figure 4.2. Illustration of Lemma 4.2. Onedimensional subfilters of quincunx filters are given by summation along the diagonals.

4.2.3. Coiflets

So far, we have focussed ourselves to design wavelets with some degree of regularity and a certain number of vanishing moments, which are closely related to regularity issues. But for (numerical) applications it is often useful to have also some vanishing (shifted) moments for the scaling functions to get e.g. reliable one-point quadrature formulae from wavelet representations [20] [161], see also Proposition 5.3. Such wavelets were first constructed by I. DAUBECHIES [63] for one dimension and named *coiflets* after R. R. COIFMAN who was the one to suggest designing such a wavelet family. In this paragraph, we will give a generalization by building nonseparable coiflets in multiple dimensions. We start by formally describing the general case, i.e. giving the mathematical characterization of multidimensional coiflets.

DEFINITION* 4.7 Suppose $\{\varphi, \psi^1, \dots, \psi^{|q|-1}\} : \mathbb{R}^n \longrightarrow \mathbb{R}$ is a family of a compactly supported scaling function and $|q|-1$ associated wavelets in \mathbb{R}^n . If

$$\mu_{\vec{m}}^{\psi^j} = 0 \quad \forall j = 1, \dots, |q|-1 \quad \text{and} \quad (4.12)$$

$$\mu_{\vec{m}, \vec{s}}^{\varphi} = \delta_{\vec{m}} \quad (4.13)$$

for a fixed $\vec{s} \in \mathbb{R}^n$ and all exponents $\vec{m} \in \mathbb{N}_0^n$ with $|\vec{m}| \leq \nu \in \mathbb{N}$ holds true, then $\{\varphi, \psi^1, \dots, \psi^k\}$ is called a coiflet family of order ν with moment shift \vec{s} .³

By a straightforward calculation, the vanishing shifted moments condition (4.13) may also be given in the filter domain:

COROLLARY* 4.8 Relation (4.13) is equivalent to $\sum_{\vec{j} \in \mathbb{Z}^n} \vec{j}^{\vec{m}} \cdot h_{\vec{j}} = \vec{s}^{\vec{m}}$.

While equations (4.12) are the usual vanishing moment conditions for the wavelet(s), whose fulfillment may be guaranteed by the choice of the basic filters, the relations in Corollary 4.8 must be implemented afterwards and lead to a set of linear equations in the coefficients of the basic filters.

EXAMPLE* 4 Consider the task of building a nonseparable twodimensional quincunx coiflet filter of e.g. first order and with moment shift $\vec{s} = \begin{pmatrix} 2 \\ 2 \end{pmatrix}$. Again, the fundamental filters h_1, h_2, h_3, h_4 and g_1, g_2 of Paragraph 4.2.2 may be chosen to build a suited family of basic filters. Additionally, we know the optimal shape of such a filter from Example 3 and since onedimensional coiflet filters with integer shift $s = 2$ are well-known [63], we also may apply Lemma 4.2 within the filter design procedure. One obtains a set of solutions, from which we quote the following solution, which corresponds to the closest-to-symmetry coiflet filter with the claimed properties.

$$h_{j,k} = \frac{1}{256} \cdot \begin{bmatrix} & -7 - 5\sqrt{7} & -7 + \sqrt{7} & & & \\ & 14 - 2\sqrt{7} & 42 + 6\sqrt{7} & 14 - 2\sqrt{7} & -14 + 6\sqrt{7} & \\ 1 - \sqrt{7} & 5 + \sqrt{7} & 84 + 28\sqrt{7} & 84 - 28\sqrt{7} & 1 - \sqrt{7} & -3 + \sqrt{7} \\ & 14 - 10\sqrt{7} & 42 + 6\sqrt{7} & 14 - 10\sqrt{7} & -14 + 6\sqrt{7} & \\ & -7 + 3\sqrt{7} & -7 + \sqrt{7} & & & \end{bmatrix}$$

Applying Theorem 4.6, we find that the corresponding scaling function and wavelet are (as their onedimensional counterparts) continuous but not differentiable functions $\varphi, \psi : \mathbb{R} \rightarrow \mathbb{R}^2$. A plot of $\varphi(\vec{x})$ is given in Figure 4.3.

4.2.4. Filter Optimization

Die Möglichkeit ist stets größer als die Wirklichkeit.

—ERNST BLOCH

³In the original coiflet construction [63] only integer moment shifts were considered, while we use the more general case by allowing real valued shifts as it was proposed in [204] [161].

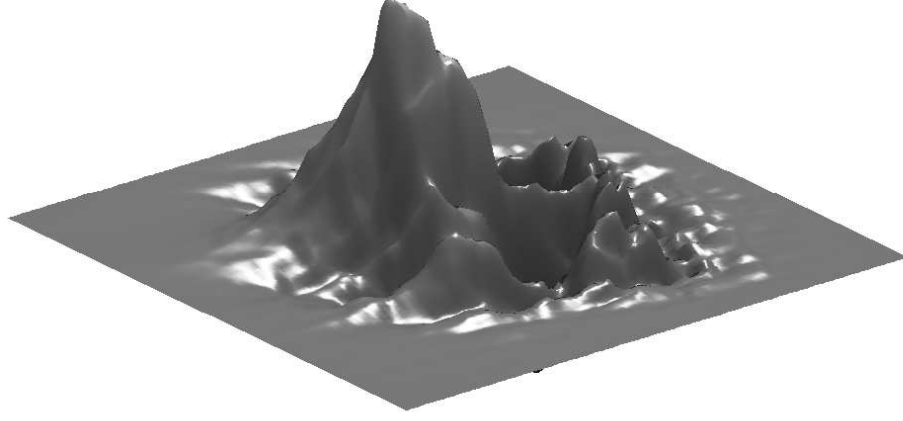


Figure 4.3. The scaling function for Example 4.

It was already shown that nonseparable filter design yields some more degrees of freedom in the design procedures than it is the case with separable filters. In this paragraph, we will show some examples, how these degrees of freedom might be employed to optimize the filters with respect to certain criteria. In particular, we will give examples that show how an optimized frequency resolution might be achieved, how close-to-isotropy solutions may be found and finally, we will design a filter which leads to a scaling function with a possibly high smoothness value.

Optimized frequency resolution. The scaling filter $H(\vec{\omega})$ of a MRA might always be interpreted as a lowpass filter. Therefore, it is a desirable property to have high energy compaction in the lowpass band of the filter, that is, a preferably large part of the signal information shall be preserved by lowpass filtering. Hereby, we want the lowpass band $L \subset [-\pi, \pi]^n$ to be defined by the set of all frequencies $\vec{\omega}$ with

$$L = \{ \vec{\omega} \mid \|\vec{\omega}\| \leq \rho \}, \quad \rho \in \mathbb{R}_{>0},$$

where $\|\cdot\|$ denotes some suited norm. The filter optimization now consists in solving the maximization problem

$$E_H(L) = \int_L |H(\vec{\omega})|^2 d\vec{\omega} = \max! . \quad (4.14)$$

We consider two different examples in 2D. In the first case, we take the EUCLIDEAN 1-norm and $\rho = \pi$. Then $L = \{ \vec{\omega} \mid |\omega_1| + |\omega_2| \leq \pi \}$, which means

that the lowpass band coincides with the shape of an ideal quincunx filter. We obtain

$$\begin{aligned}
E_H(L) &= \int_{|\omega_1|+|\omega_2|\leq\pi} |H(\vec{\omega})|^2 d\vec{\omega} \\
&= \int_{|\omega_1|+|\omega_2|\leq\pi} \sum_{\vec{k}\in\mathbb{Z}^2} f_{k_1,k_2} \cdot \cos(k_1\omega_1 + k_2\omega_2) d\vec{\omega} \\
&= \int_{[-\frac{\pi}{2},\frac{\pi}{2}]^2} \sum_{\vec{k}\in\mathbb{Z}^2} f_{k_1,k_2} \cdot \cos(k_1(\omega_1 + \omega_2) + k_2(\omega_1 - \omega_2)) d\vec{\omega} \\
&= \sum_{\vec{k}\in\mathbb{Z}^2} f_{k_1,k_2} \cdot \int_{[-\frac{\pi}{2},\frac{\pi}{2}]^2} \cos(k_1(\omega_1 + \omega_2) + k_2(\omega_1 - \omega_2)) d\vec{\omega}.
\end{aligned}$$

Since the coefficients f_{k_1,k_2} are polynomials of degree 2 in the filter coefficients h_{k_1,k_2} , the same is true for the lowpass energy functional $E_H(L)$. Now the optimization of the lowpass energy is simply done by taking the partial derivatives with respect to the filter coefficients h_{k_1,k_2} and solving the resulting linear equation system with respect to these filter coefficients.

In our second example, we choose to consider the EUCLIDEAN 2-norm and assume $0 < \rho^2 \leq \pi^2$. This choice corresponds to a lowpass band with radially isotropic frequency resolution. Now, the energy functional reads

$$\begin{aligned}
E_H(L) &= \int_{|\omega_1|^2+|\omega_2|^2\leq\rho^2} |H(\vec{\omega})|^2 d\vec{\omega} \\
&= \int_{|\omega_1|^2+|\omega_2|^2\leq\rho^2} \sum_{\vec{k}\in\mathbb{Z}^2} f_{k_1,k_2} \cdot \cos(k_1\omega_1 + k_2\omega_2) d\vec{\omega}.
\end{aligned}$$

Replacing the frequency coordinates (ω_1, ω_2) by polar coordinates (θ, ω) and using the abbreviation $\kappa = \arctan \frac{n_1}{n_2}$, this expression becomes

$$\begin{aligned}
E_H(L) &= \int_0^\rho \int_0^{2\pi} \sum_{\vec{k}\in\mathbb{Z}^2} f_{k_1,k_2} \cdot \omega \cdot \cos(\omega\sqrt{k_1^2 + k_2^2} \cdot \sin(\kappa - \theta)) d\theta d\omega \\
&= 2\pi \cdot \sum_{\vec{k}\in\mathbb{Z}^2} f_{k_1,k_2} \cdot \int_0^\rho \omega \cdot J_0(\omega\sqrt{k_1^2 + k_2^2}) d\omega \\
&= 2\pi\rho \cdot \sum_{\vec{k}\in\mathbb{Z}^2} f_{k_1,k_2} \cdot \frac{J_1(\rho\sqrt{k_1^2 + k_2^2})}{\sqrt{k_1^2 + k_2^2}}.
\end{aligned}$$

Hereby, $J_m(\cdot)$ denotes the BESSEL function of first kind of order m ; the BESSEL function values can be evaluated numerically with arbitrary precision. Thus,

the energy functional is again given by a polynomial of degree two in the filter coefficients and the optimization may be done as in the first case. A completely different approach to optimized frequency resolution is the *splitting trick* [46], which targets at dividing the frequency bands into narrower passbands.

Close-to-isotropy solutions. A filter with filter coefficients $h_{\vec{k}}$ is called *linear phase* if the phase of the transfer function $H(\vec{\omega})$ is a linear function in $\vec{\omega}$, that is for a suited $\vec{l} \in \frac{1}{2}\mathbb{Z}^n$ we have

$$H(\vec{\omega}) = |H(\vec{\omega})| \cdot e^{-i\langle \vec{l}, \vec{\omega} \rangle}.$$

It is easy to prove that the linear phase property is equivalent to have a symmetric filter with respect to the point \vec{l} and this consequently also gives a symmetric scaling function. However, it is a known fact, that nontrivial orthogonal wavelets with compact support can never be fully symmetric with respect to the axes (and especially not radially symmetric) and therefore, it might be necessary to find filters that lead to *least asymmetric* scaling functions. Given a *symmetry center* \vec{l} this problem is solved by minimizing the functional

$$E_H(\vec{l}) = \sum_{\vec{k} \in \mathbb{Z}^n, \vec{\sigma} \in S_n(\{-1,1\}), \vec{j} \in S_n(\vec{k})} (h_{\vec{l}+\vec{k}} - h_{\vec{l}+\vec{\sigma}\vec{j}})^2,$$

which gives a symmetry distortion measure for the filter coefficients. As in the frequency optimization procedure, this is a quadratic functional and the minimization essentially reduces to the solution of a set of linear equations in the filter coefficients.⁴

EXAMPLE* 5 We shall try to find the least asymmetric 4×4 filter for the regular sampling matrix $\mathbf{Q} = 2 \cdot \mathbf{Id}_2$ satisfying at least the sum rules of first order. Such a filter might be seen as a generalization of the DAUBECHIES tap-4 filter in 1D. Obviously, this filter consists of 16 taps and it turns out, that only 14 conditions are to be fulfilled to create such a filter (4 orthogonality constraints, 9 sum rules and one for normalization). The remaining two degrees of freedom are utilized to reduce the asymmetry of the filter as much as possible. We choose the center as $\vec{l} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ and obtain the solution

⁴In 1D, this procedure may be done in a more elegant way: the onedimensional transfer function $H(\omega)$ can be decomposed into linear factors and by simple calculus, the phase function $p(\omega)$ can be explicitly evaluated in this case. However, such factorizations do not exist in higher dimensions and therefore, this method cannot applied in the present situation.

$$h_{j,k} = \frac{1}{96} \cdot \begin{bmatrix} 3 + 6\sqrt{3} & 11 + 6\sqrt{3} - \sqrt{10} & 3 & -5 + \sqrt{10} \\ 11 + 6\sqrt{3} - \sqrt{10} & 17 + 6\sqrt{3} + 2\sqrt{10} & 7 + \sqrt{10} & 1 - 2\sqrt{10} \\ 3 & 7 + \sqrt{10} & 15 - 6\sqrt{3} & 11 - 6\sqrt{3} - \sqrt{10} \\ -5 + \sqrt{10} & 1 - 2\sqrt{10} & 11 - 6\sqrt{3} - \sqrt{10} & 5 - 6\sqrt{3} + 2\sqrt{10} \end{bmatrix}$$

which is symmetric with respect to the diagonal $k_1 = k_2$ and least point-asymmetric with respect to the point $\vec{l} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

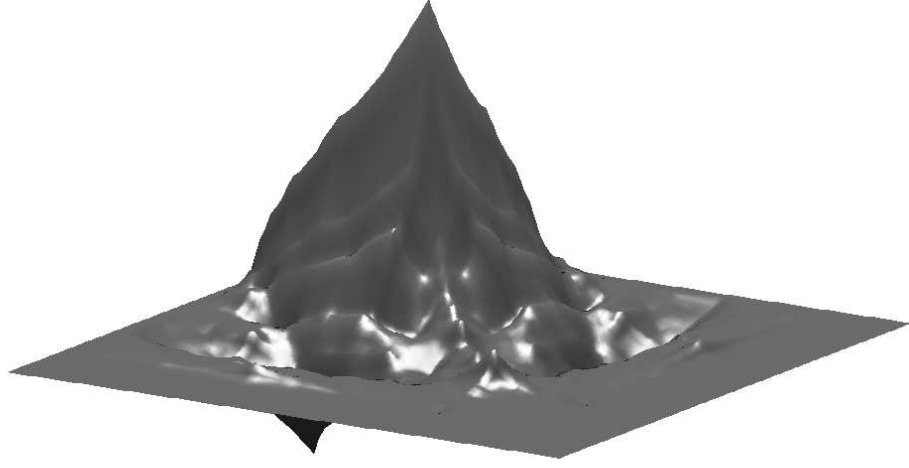


Figure 4.4. The scaling function for Example 5.

Remark. Using *polynomial connection coefficients*, which will be introduced later in this thesis, is another possibility to design filters that yield scaling functions optimally close to (radial) symmetric functions.

Possibly high smoothness. For visual representation, we have already learned that it is a good property to have basis (or more generally: representing) functions with some degree of smoothness [8] [245]. On the other hand, for the important quincunx sampling scheme, not even one *nice* nonseparable orthogonal scaling functions with one time continuous derivatives is known till today [47] [228].⁵ Using all the filter design techniques we have presented in this chapter, we have found the following scaling filter after searching the whole

⁵This is not exactly true. There exist two design methods to find nonseparable filters that might give smooth orthogonal scaling functions by perturbing 1D filters [17] [11]. However, the resulting functions are close to separable solutions and do not possess any desirable properties to apply them in image processing.

space of possible orthogonal filters with the same shape having at least two vanishing moments. By minimizing the largest eigenvalues of the operators \mathbf{T}_1 and \mathbf{T}_2 (see Theorem 4.6), we obtain the filter

$$\begin{aligned}
 h_{4,6} &= \frac{835732 - 64636\alpha + 2521\beta - 52\alpha\beta}{4736000} \\
 h_{5,6} &= \frac{1671464 - 129272\alpha + 5042\beta - 104\alpha\beta}{4736000} \\
 h_{3,5} &= \frac{336248 - 20904\alpha - 6880\beta + 312\alpha\beta}{4736000} \\
 h_{4,5} &= \frac{308251 + 10452\alpha - 5420\beta - 156\alpha\beta}{4736000} \\
 h_{5,5} &= \frac{439016 + 512\alpha - 3330\beta}{4736000} \\
 h_{6,5} &= \frac{878032 + 1024\alpha - 6660\beta}{4736000} \\
 h_{2,4} &= \frac{1584068 + 64636\alpha + 557\beta + 52\alpha\beta}{4736000} \\
 h_{3,4} &= \frac{2391136 + 129272\alpha + 6904\beta + 104\alpha\beta}{4736000} \\
 h_{4,4} &= \frac{1067984 - 44032\alpha + 11840\beta}{4736000} \\
 h_{5,4} &= \frac{-850342 + 22016\alpha + 1480\beta}{4736000} \\
 h_{6,4} &= \frac{-114428 - 6916\alpha + 809\beta + 52\alpha\beta}{4736000} \\
 h_{7,4} &= \frac{-228856 - 13832\alpha + 1618\beta + 104\alpha\beta}{4736000} \\
 h_{1,3} &= \frac{287202 + 20904\alpha - 5448\beta - 312\alpha\beta}{4736000} \\
 h_{2,3} &= \frac{-143601\alpha - 10452\beta + 156\alpha\beta}{4736000} \\
 h_{3,3} &= \frac{1760264 - 512\alpha + 1110\beta}{4736000} \\
 h_{4,3} &= \frac{308928 - 1024\alpha - 8880\beta}{4736000} \\
 h_{5,3} &= \frac{158648 + 17576\alpha - 4960\beta - 312\alpha\beta}{4736000} \\
 h_{6,3} &= \frac{-239349 - 8788\alpha + 3940\beta + 156\alpha\beta}{4736000} \\
 h_{2,2} &= \frac{-1953764 + 44032\alpha + 8880\beta}{4736000} \\
 h_{3,2} &= \frac{976882 - 22016\alpha - 4440\beta}{4736000} \\
 h_{4,2} &= \frac{231348 + 6916\alpha - 1667\beta - 52\alpha\beta}{4736000} \\
 h_{5,2} &= \frac{-284704 + 13832\alpha + 1976\beta - 104\alpha\beta}{4736000}
 \end{aligned}$$

$$\begin{aligned} h_{3,1} &= \frac{103682 - 17576\alpha - 3432\beta + 312\alpha\beta}{4736000} \\ h_{4,1} &= \frac{-51841 + 8788\alpha + 1716\beta - 156\alpha\beta}{4736000} \end{aligned}$$

with the coefficient arrangement

$$h = \begin{bmatrix} & & & h_{4,6} & h_{5,6} & & \\ & & & h_{4,5} & h_{5,5} & h_{6,5} & \\ & h_{2,4} & h_{3,4} & h_{4,4} & h_{5,4} & h_{6,4} & h_{7,5} \\ h_{1,3} & h_{2,3} & h_{3,3} & h_{4,3} & h_{5,3} & h_{6,3} & \\ & h_{2,2} & h_{3,2} & h_{4,2} & h_{5,2} & & \\ & & h_{3,1} & h_{4,1} & & & \end{bmatrix}.$$

The choices $\alpha = \sqrt{239}$ and $\beta = \sqrt{13135}$ give in fact a continuously differentiable scaling function for the quincunx dilation matrix $\mathbf{Q}_{\text{quin},2}$. This is the first algebraically given filter that leads to a \mathcal{C}^1 function for this scaling matrix. Its HÖLDER exponent can be estimated to attain at least the value 1.04 and we conjecture $1.0757 \dots$ to be the true value, which is the upper bound given by the largest eigenvalue. The estimations were found by applying the techniques developed in [95] to Theorem 4.6. A plot of the related scaling function is given in Picture 4.5.

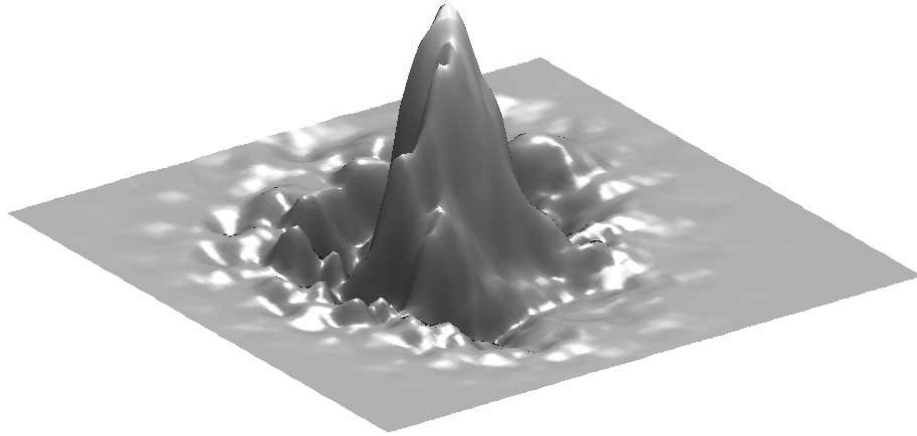


Figure 4.5. The first orthogonal \mathcal{C}^1 scaling function for the quincunx dilation matrix $\mathbf{Q}_{\text{quin},2}$.

The methods developed in this Chapter make it possible to design nonseparable wavelets with certain features. But we shall also remark that they do not help to design infinite function families of increasing smoothness which

generalizes the famous construction of DAUBECHIES [60]. Although several researchers tried to attack this problem over the last 10 years, a positive answer to this question seems to be still out of reach.

4.2.5. Other Approaches

For completeness and comparability reasons, we will briefly review and comment some other approaches to multidimensional wavelet filter design now.

Maybe the most popular method for multidimensional (wavelet) filter design is the usage of *polyphase lattice structures* as developed in [224] [223]. It goes back to the work of VAIDYANATHAN and his co-workers. The main idea of this method depends on the polyphase representation of a set of filters with a perfect reconstruction property by a product of paraunitary matrices. In particular, this means that the polyphase matrix can be written as

$$\mathbf{H} = [H_{ij}(\vec{\omega})]_{ij} = \mathbf{R}_{\alpha_0, \xi_0, \eta_0} \cdot \prod_{k=1}^v \mathbf{J}_{\vec{m}_k, \chi_k} \cdot \mathbf{R}_{\alpha_k, \xi_k, \eta_k}, \quad (4.15)$$

where $\mathbf{R}_{\alpha_k, \xi_k, \eta_k}$ denotes a GIVENS rotation about the angle α_k with respect to the axes ξ_k, η_k and

$$\mathbf{J}_{\vec{m}_k, \chi_k} = \mathbf{Id}_n + (e^{i \cdot \langle \vec{\omega}, \vec{m}_k \rangle} - 1) \cdot \vec{e}_{\chi_k} \cdot \vec{e}_{\chi_k}^T$$

are the elementary paraunitary building blocks (\vec{e}_j denotes any unit vector in \mathbb{R}^n). Due to the orthogonality of GIVENS rotations, this obviously gives a set of orthogonal filters by construction. This so-called *lattice factorization* attracted many researchers working with multidimensional filter banks, since it could be implemented in a way that it automatically generates sets of orthogonal (or optionally linear phase) filters, which is in contrast to the *more difficult* orthogonality constraints given by the quadratic equations (4.2). But this easy-to-achieve orthogonality has to be paid by the price of significantly harder conditions to obtain filters with some vanishing moments. Using the method described in the previous sections, vanishing moments could be established by solving a set of linear equations, while in the case of paraunitary lattice factorizations, this becomes a set of nonlinear equations of degree v . Clearly, especially for large values of v (bigger filters) the total complexity for the design procedure is thus much higher than with our method. Another disadvantage of lattice factorization is that due to the non-commutativity of matrices, there exists no unique ordered factorization for (4.15) and it is not a priori clear which factorization should be ideally used to implement desired features into a filter bank. Finally, as we shall see in Section 4.4, lattice structures are conceptually unsuited to build wavelet frames, which do not fit into the polyphase framework, since they yield overdetermined representations.

In [142], MAASS constructs dyadic orthogonal twodimensional wavelets by describing the set H_2 of all scaling coefficients satisfying the orthogonality condition as an implicit function. By computing the tangent space $T \subset H_2$ of onedimensional scaling coefficients in $H_1 \subset H_2$, nonseparable twodimensional scaling filters are obtained. Mathematically, this may be formulated as follows.

LEMMA 4.9 *Let $H(\omega) \in H_1$ denote the transfer function of the scaling filter $\{h_k \mid k = 0, \dots, 2n - 1\}$ for a dyadic orthogonal 1D wavelet. Denote the tangent vector by $G(\omega) \in T$. Let $P(\omega_1, \omega_2), Q(\omega_1, \omega_2)$ be any (π, π) -periodic trigonometric polynomials satisfying*

$$|P(\omega_1, \omega_2)|^2 + |Q(\omega_1, \omega_2)|^2 \equiv 1, \quad P(0, 0) = 1.$$

Then

$$\tilde{H}(\omega_1, \omega_2) = P(\omega_1, \omega_2) \cdot H(\omega_1) + Q(\omega_1, \omega_2) \cdot G(\omega_1)$$

satisfies the 2D orthogonality relation for the quincunx dilation matrix.

Proof. See [142]. □.

This construction can be easily modified to other dilations than quincunx and also to higher dimensions. However, as was stated in [140], it was not possible to derive filters that lead to differentiable wavelets or to find closed forms for filters with a higher number of vanishing moments via this method. Additionally, the *tangent space* construction leads to rather large filters since the trigonometric polynomials P, Q cannot be chosen arbitrary but have to satisfy a criterion which increases their number of coefficients and therefore also the filter size. Also the integration of other properties into the filters is more complicated, since it has to consider the constraint on P and Q . Except the very descriptive geometric construction, this method does not help to overcome the problems in multidimensional filter design.

A more theoretical approach to multidimensional orthogonal filter design was made by HOLSCHNEIDER and PINKALL in [112]. They showed that the transfer function of a quadrature mirror filter associated to a grid $\Gamma \subset \mathbb{Z}^n$ can be identified with the group of polynomial loops $\{f : \mathbb{T}^n \rightarrow \mathcal{U}(|\mathbb{Z}^n/\Gamma|) : f \in \Pi_k\}$ (\mathbb{T}^n denotes the n -unit sphere and $\mathcal{U}(k)$ the set of unitary $k \times k$ matrices). Moreover, they could also prove that QMFs with vanishing moments can be identified with cosets of certain subgroups and therefore, the multidimensional filter design problem can be shown to be equivalent to the problem of factoring all polynomial loops. However, since this problem is unsolved for more than one dimension, this only yields another formulation of the original problem and does not give rise to a general filter design procedure as long as the factorization

problem remains open. On the other hand, in [113] some very specific new filter examples were obtained using this method.

The presented list of approaches to multidimensional filter design is of course not exhaustive, but it demonstrated the specific difficulties from different points of view and stressed the advantages of the methods presented in this thesis. Several more approaches to the problem are thinkable. One such possibility might be an algebraic geometry method by considering the orthogonality constraints and other design equations (some vanishing moments, close-to-isotropic solutions, ...) as varieties in a suited highdimensional space and trying to specify intersection points of these algebraic curves. Unfortunately, this is again an unsolved problem for dimensions greater than one, but gives again another mathematical characterization, which might be helpful in future approaches.

The author also tried to attack the problem by maintaining so-called *hypercomplex* FOURIER transforms [27] [168] in the definition of the transfer function associated to a multidimensional filter in order to circumvent the non-validity of the FEJER-RIESZ Theorem 2.10. But since hypercomplex number systems do not have an algebraic field structure, we were not successful in the trial to generalize the result of FEJER and RIESZ to algebraic structures of this kind.

4.3. Biorthogonal Wavelets in \mathbb{R}^n

Symmetry, as wide or as narrow as you may define its meaning, is one idea by which man through the ages has tried to comprehend and create order, beauty, and perfection.

—HERMANN WEYL [241]

Orthogonality for wavelets is a strict requirement, which involves some advantages (basis property, stable inversions) but also some deficiencies as non-symmetry or rather low smoothness for the scaling functions and wavelets⁶ as well as the hard problem of the filter design itself. Mathematically, orthogonality does mean to use the same bases for the analysis as well as for the synthesis operation in a signal processing task. Giving up this restriction and allowing different bases for the analysis and synthesis but preserving the stable reconstruction property, we arrive at *biorthogonal* wavelets which were originally introduced in [45]. These are given by a pair of two scaling filters H, G

⁶A onedimensional orthogonal filter of length $2n$ is known to give approximately a HÖLDER regularity of only $\alpha_n \approx 0.2075n$ [60], while a B-spline of the same length achieves the smoothness order $2n - 2 - \epsilon$.

satisfying the *biorthogonality relation*

$$\sum_{j=1}^{|q|} H(\vec{\omega} + 2 \cdot \pi \cdot \mathbf{Q}^{-T} \cdot \vec{r}_j) \cdot \overline{G(\vec{\omega} + 2 \cdot \pi \cdot \mathbf{Q}^{-T} \cdot \vec{r}_j)} \equiv 1 \quad \text{a.e.}, \quad (4.16)$$

which is the biorthogonal generalization of (4.1). The design of biorthogonal multidimensional filters is much simpler than in the orthogonal case, because it is possible to choose *any* desired filter as a *primal filter* H and one directly obtains the according *dual filter* G by solving (4.16) which are just linear equations in the filter coefficients; vanishing moments or other desired properties can be implemented analogously to the orthogonal case by solving another set of linear equations.

Another possibility for biorthogonal filter design is the application of the MCCLELLAN transform. Given a pair H_1, G_1 of linear-phase onedimensional dual scaling filters written as

$$\{H_1/G_1\}(\omega) = \sum_{k=-j}^j \{h^1/g^1\}_k \cdot \cos(k\omega) = \sum_{k=-j}^j \{h^1/g^1\}_k \cdot T_k(\cos(k\omega)),$$

where T_k denotes the k -th TCHEBYCHEFF polynomial, a pair of multidimensional biorthogonal filters is given by

$$\{H/G\}(\vec{\omega}) = \sum_{k=-j}^j \{h^1/g^1\}_k \cdot T_k(p(\vec{\omega}))$$

if $p(\vec{\omega})$ is the transfer function of an arbitrary multidimensional filter with linear phase. Usually, the argument $p(\vec{\omega})$ of the MCCLELLAN transform is chosen to implement vanishing moments into the new filters, since these are preserved by construction, as one easily verifies. The fact that the new filter pair also satisfies (4.16) can be proven by simple evaluation and application of the fact that TCHEBYCHEFF polynomials form an orthonormal system.

There are more approaches to biorthogonal filter design. The most widely known (and maybe also the most comfortable) one among these is the lifting scheme [216], which was briefly presented at the end of the previous chapter. While lifting allows some interesting generalizations to wavelets on manifolds or more general algebraic structures (e.g. CLIFFORD algebras [160]), it does not give any new results for (bi)orthogonal wavelets on \mathbb{R}^n which are considered in this thesis.

We may conclude that biorthogonal multidimensional wavelets are much easier obtainable and can be equipped with nice features such as symmetry or higher regularity than orthogonal wavelets. On the other hand, especially establishing regularity comes along with dramatically increasing filter sizes. In [47], a dual pair of biorthogonal quincunx wavelets such that both primal and dual functions are at least once continuously differentiable were constructed. While the primal filter of this example consists of 13 taps, its dual counterpart has a size of more than 1350 taps which seems to be very impracticable. The reader may compare this size to the comparably small 24-tap orthogonal filter giving a \mathcal{C}^1 scaling function presented in the previous section. Biorthogonal wavelets will not play any further role in this thesis, since we will now introduce another theory which allows us to join all the advantages of orthogonal wavelets, biorthogonal wavelets and multidimensional splines into one mathematical concept.

4.4. Tight Wavelet Frames

So far we have only considered bases, i.e. complete function systems for signal processing and representation purposes in this thesis. But for several reasons, it might also be of interest to have an overcomplete (i.e. redundant) function system at hand. Overcompleteness is mostly helpful to stabilize ill-posed numerical problems or to better suppress the effects caused by noise in data measurements [164]; both are typical problems occurring in computer vision. The requirement for function systems with such capabilities leads us to the theory of *wavelet frames* (sometimes also called *skew structures*). It will turn out, that the allowance of overdeterminant systems will have the nice by-effect that properties like higher order regularity for relatively small filters, vanishing moments and symmetry are no longer exclusive, even if perfect reconstruction is requested.

4.4.1. Introduction

As already mentioned in Chapter 1, frames were introduced about 50 years ago in [79] developing the theory of nonharmonic FOURIER series. To start with our investigations, we choose the following more recent definition of a frame (see also [109]).

DEFINITION 4.10 *A family of functions $(\varphi_k)_{k \in I}$ in a HILBERT space V is called a frame if there exist positive constants $\infty > C_u \geq C_l > 0$ such that for all $f \in V$*

$$C_l \cdot \|f\|^2 \leq \sum_{k \in I} |\langle f, \varphi_k \rangle|^2 \leq C_u \cdot \|f\|^2 \quad (4.17)$$

holds true. If the frame bounds C_u, C_l are equal, the frame is called tight.

For a tight frame, we have $C_u = C_l$ and thus equality in the defining equation (4.17). Using this fact and applying the polarization equation, we obtain a resolution of identity

$$f = \frac{1}{C_u} \cdot \sum_{k \in I} \langle f, \varphi_k \rangle \cdot \varphi_k, \quad (4.18)$$

which holds almost everywhere except (possibly) on a set of zero measure.⁷ If the frame is built up by dilated and translated versions of a set of $m + 1$ fundamental functions ψ^i that satisfy a scaling equation like (3.7), i.e.

$$\{(\varphi_k)_{k \in I}\} = \{|\det \mathbf{Q}|^{-j/2} \psi^i(\mathbf{Q}^{-j} \cdot -\vec{k}) \mid j \in \mathbb{Z}, \vec{k} \in \mathbb{Z}^n, i \in \{0, \dots, m\}\}$$

the frame is called a *wavelet frame* and analogously a *tight wavelet frame* if $C_u = C_l$. In a series of papers, RON, SHEN and GRÖCHENIG developed the mathematical theory of (tight) wavelet frames whose fundamental construction principle we will now recall [192] [193] [194] [195] [98].

THEOREM 4.11 *As in Lemma 4.1, let \mathbf{Q} be an integer dilation matrix and $\{\vec{0} = \vec{r}_1, \vec{r}_2, \dots, \vec{r}_{|q|}\}$ be a complete set of representatives of the cosets of the reciprocal sampling lattice $\Gamma = \mathbf{Q}^T \cdot \mathbb{Z}^n$ in \mathbb{Z}^n . Additionally, let ψ^0 be a scaling function that is refinable with respect to \mathbf{Q} associated to a symbol function $H_0(\vec{\omega})$. If there exist m measurable symbol functions $(H_i(\vec{\omega}))_{i=1, \dots, m}$ satisfying*

$$\sum_{j=0}^m H_j(\vec{\omega}) \cdot \overline{H_j(\vec{\omega} + 2 \cdot \pi \cdot \mathbf{Q}^{-T} \cdot \vec{r}_k)} = \delta_{1,k}, \quad k = 1, \dots, |q|, \quad (4.19)$$

then the wavelets defined via

$$\hat{\psi}^i(\vec{\omega}) = H_i(\mathbf{Q}^{-T} \cdot \vec{\omega}) \cdot \hat{\psi}^0(\mathbf{Q}^{-T} \cdot \vec{\omega}), \quad i = 0, 1, \dots, m,$$

form a tight wavelet frame for the $L_2(\mathbb{R}^n)$ given by

$$\Psi = \left\{ |\det \mathbf{Q}|^{-j/2} \psi^i(\mathbf{Q}^{-j} \cdot -\vec{k}) \mid j \in \mathbb{Z}, \vec{k} \in \mathbb{Z}^n, i \in \{0, \dots, m\} \right\}.$$

The frame bound for Ψ is $C_u = C_l = 1$.

Proof. This theorem summarizes several results from [193] and [194]. We refer the interested reader to these articles. \square .

⁷For $C_u = C_l$, (4.18) looks similar to the classical orthogonal wavelet expansion, however, tight frames are generally not orthogonal bases, even for $C_u = C_l = 1$. It can be easily shown that this is only the case for tight frames with frame bound 1 and under the additional condition that $\|\varphi_k\| = 1$ for all $k \in I$ [61].

The condition (4.19) looks very much alike the orthogonality condition (4.1) at a first glance. And there is in fact some similarity: RON and SHEN established a *unitary extension principle* for the design of tight wavelet frames [195], which can be viewed as a generalization of LAWTON's construction of orthonormal wavelets in [134]. However, there is a fundamental difference; while in orthogonal wavelet constructions, the number of basic wavelets is strictly limited to $|q| - 1$, for the construction of tight wavelet frames one is allowed to choose $m > |q| - 1$ wavelets to supplement the scaling function in order to build up a frame. This gives additional degrees of freedom and allows the abovementioned improvements in the filter design procedure. But the direct design of tight wavelet frames is in general as hard as the design of orthogonal wavelets, since it again comes essentially down to the problem of solving a set of nonlinear equations. This can be directly done for small filters⁸ and becomes unfeasible for very large filters. For this reason, in [194] an algorithm was developed to build bigger tight wavelet frames with (possibly) more vanishing moments and increasing smoothness from smaller prototypes which can be found algebraically. For this simplified construction, one has to pay the price that the number m of required wavelets increases by a constant in each step; the area of the support of the new frame functions even increases much faster, because these are essentially given by

$$\varphi_{\text{new}} = |q|^{-1} \cdot \psi^i(\vec{\omega}) * \phi^j(\mathbf{Q}^{-1} \cdot \vec{\omega}),$$

where ϕ^j is some suited smoothness and/or vanishing moments increasing distribution.

4.4.2. A New and Easy Way to Build Tight Wavelet Frames

Our plan for the rest of this section is to develop an alternative construction method for tight wavelet frames derived from simple prototypes. This new method will have the advantage that the support area of the new functions will be significantly smaller than in the construction [194]. On the other hand we have to accept (at least in the most cases) the drawback of higher values for m . Before we can state this result, we need to evolve some polynomial identity, that we will use for the proof. To start with this, we introduce the n -tuple set

$$K(n, j) = \left\{ \vec{k} = (k_1, k_2, \dots, k_n), k_i \in 0, \dots, j \ \forall i \mid \sum_{i=1}^n k_i = j \right\}$$

⁸RON and SHEN used *box splines* [25] to construct such smaller basic frames. Box splines are a multidimensional generalization of B-splines and share the property of being refinable with respect to a suited scaling matrix with their onedimensional prototypes.

of all possible ordered partitions of $j \in \mathbb{N}$ into at most n nonnegative numbers $k_i \in \mathbb{N}_0$. It is a well known combinatorial fact and easy to prove, that

$$|K(n, j)| = \binom{n-1+j}{j}$$

holds true for all $n, j \in \mathbb{N}$. Equipped with this, we can prove the following.

LEMMA* 4.12 For all $j \in \mathbb{N}$ and $n \in \mathbb{N}$ the following identity holds true:

$$(x_1 + \dots + x_n)^j = \sum_{(k_1, \dots, k_n) \in K(n, j)} \left(\frac{j!}{k_1! \cdot k_2! \cdot \dots \cdot k_n!} \right) \cdot x_1^{k_1} \cdot x_2^{k_2} \cdot \dots \cdot x_n^{k_n}.$$

Proof. The claim can be easily shown by direct evaluation under repeated application of the binomial theorem:

$$\begin{aligned} (x_1 + \dots + x_n)^j &= \sum_{l_1=0}^j \binom{j}{l_1} \cdot x_1^{j-l_1} \cdot (x_2 + x_3 + \dots + x_n)^{l_1} \\ &= \sum_{l_1=0}^j \sum_{l_2=0}^{l_1} \binom{j}{l_1} \cdot \binom{l_1}{l_2} \cdot x_1^{j-l_1} \cdot x_2^{l_1-l_2} \cdot (x_3 + \dots + x_n)^{l_2}. \end{aligned}$$

After $n-1$ repeated applications of the binomial theorem, we obtain

$$\begin{aligned} &= \sum_{l_1=0}^j \dots \sum_{l_{n-1}=0}^{l_{n-2}} \binom{j}{l_1} \cdot \binom{l_1}{l_2} \cdot \dots \cdot \binom{l_{n-2}}{l_{n-1}} \cdot x_1^{j-l_1} \cdot x_2^{l_1-l_2} \cdot \dots \cdot x_{n-1}^{l_{n-2}-l_{n-1}} \cdot x_n^{l_{n-1}} \\ &= \sum_{k_1=0}^j \sum_{k_2=0}^{j-k_1} \dots \sum_{k_{n-1}=0}^{j-k_1-\dots-k_{n-2}} \frac{j! \cdot x_1^{k_1} \cdot \dots \cdot x_{n-1}^{k_{n-1}} \cdot x_n^{j-k_1-\dots-k_{n-1}}}{k_1! \cdot k_2! \cdot \dots \cdot k_{n-1}! \cdot (j-k_1-\dots-k_{n-1})!} \\ &= \sum_{(k_1, \dots, k_n) \in K(n, j)} \left(\frac{j!}{k_1! \cdot k_2! \cdot \dots \cdot k_n!} \right) \cdot x_1^{k_1} \cdot x_2^{k_2} \cdot \dots \cdot x_n^{k_n}. \end{aligned}$$

This is exactly the desired relation. \square .

Applying this lemma, we are ready to state and prove the new tight wavelet frame construction.

THEOREM* 4.13 Given a set of filters $H_i(\vec{\omega})$, $i = 0, 1, \dots, m$ that satisfy the tight wavelet frame conditions (4.19)

$$\sum_{k=0}^m H_k(\vec{\omega}) \cdot \overline{H_k(\vec{\omega} + 2 \cdot \pi \cdot \mathbf{Q}^{-T} \cdot \vec{r}_l)} = \delta_{1,l}, \quad l = 1, \dots, |q|.$$

As before, Γ denotes the complete set of representatives of the reciprocal lattice $\mathbb{Z}^n / \mathbf{Q}^T \cdot \mathbb{Z}^n$, where \mathbf{Q} is the related scaling matrix. Then, for any positive integer $j \in \mathbb{N}$, the $\binom{m+1}{j}$ filters

$$G_{(k_0, \dots, k_m)}(\vec{\omega}) = \left(\frac{j!}{k_0! \cdot k_1! \cdots k_m!} \right)^{1/2} \cdot \prod_{i=0}^m H_i(\vec{\omega})^{k_i}$$

with $(k_0, \dots, k_m) \in K(m+1, j)$ also satisfy the frame conditions

$$\sum_{\vec{k}=(k_0, \dots, k_m) \in K(m+1, j)} G_{\vec{k}}(\vec{\omega}) \cdot \overline{G_{\vec{k}}(\vec{\omega} + 2 \cdot \pi \cdot \mathbf{Q}^{-T} \cdot \vec{r}_l)} = \delta_{1,l}, \quad l = 1, \dots, |q|$$

and thus generate a new tight wavelet frame for the space $L_2(\mathbb{R}^n)$.

Proof. With the result of Lemma 4.12, we can do the following straightforward computation, which remains obviously valid for all $l = 1, \dots, |q|$.

$$\begin{aligned} & \sum_{\vec{k}=(k_0, \dots, k_m) \in K(m+1, j)} G_{\vec{k}}(\vec{\omega}) \cdot \overline{G_{\vec{k}}(\vec{\omega} + 2 \cdot \pi \cdot \mathbf{Q}^{-T} \cdot \vec{r}_l)} \\ &= \sum_{(k_0, \dots, k_m) \in K(m+1, j)} \left(\frac{j!}{k_0! \cdot k_1! \cdots k_m!} \right) \cdot \prod_{i=0}^m H_i(\vec{\omega})^{k_i} \cdot \overline{H_i(\vec{\omega} + 2 \cdot \pi \cdot \mathbf{Q}^{-T} \cdot \vec{r}_l)^{k_i}} \\ &= \left(\sum_{k=0}^m H_k(\vec{\omega}) \cdot \overline{H_k(\vec{\omega} + 2 \cdot \pi \cdot \mathbf{Q}^{-T} \cdot \vec{r}_l)} \right)^j = \delta_{1,l}^j = \delta_{1,l}. \end{aligned}$$

And this completes the proof of Theorem 4.13. \square .

We should add some remarks to this theorem. First, it seems to be disadvantageous, that the number of involved filters increases very fast with j . But since the new filters are only products of the initial filters (up to scalar multiples) in FOURIER space, the new filters can be realized by repeated convolution of the initial filters in spatial domain and thus, no additional effort is necessary to perform the filtering. This is in contrast to the construction of RON and SHEN, who obtain successively higher order tight frames by

$$\varphi_{\text{new}} = |q|^{-1} \cdot \psi^i(\vec{\omega}) * \phi^j(\mathbf{Q}^{-1} \cdot \vec{\omega}),$$

On one hand, their construction increases the number of functions only by a constant in each convolution step, but on the other hand, the support size of the frames designed this way increases $|q|$ times faster in each step by construction.

Furthermore, if one applies Theorem 4.13 to a pair of a dyadic orthogonal scaling function and wavelet, then we have $m = |q| - 1 = 1$ and thus, we only need $\binom{j+1}{j} = j + 1$ functions to build up the new frame. This means, that we can reach a lower number of functions than [194] to form a tight frame in this special case; on the other hand, these functions cannot be symmetric, which may be a drawback in certain applications. But at least partial symmetry can be achieved by the following trick:

LEMMA* 4.14 *Given a pair $\varphi(\vec{x}), \psi(\vec{x})$ of a dyadic orthogonal scaling function and wavelet with corresponding symbol functions $H_0(\vec{\omega}), H_1(\vec{\omega})$. A tight wavelet frame consisting of three functions can be built from the new symbols*

$$G_0 = H_0 \cdot \overline{H_0}, \quad G_1 = \sqrt{2} \cdot H_0 \cdot \overline{H_1} \quad \text{and} \quad G_2 = H_1 \cdot \overline{H_1}.$$

At least the functions corresponding to $G_0(\vec{\omega})$ and $G_2(\vec{\omega})$ are symmetric and antisymmetric respectively and moreover, the scaling function associated to G_0 is interpolating.

Proof. That the three functions form a frame can be shown by direct evaluation under application of Lemma 4.12 (see also the proof of Theorem 4.13). G_0 and G_2 are (anti)symmetric by construction. The scaling function is also interpolating since its symbol satisfies the interpolatory condition [72]

$$\sum_{l=1}^{|q|} G_{\vec{k}}(\vec{\omega} + 2 \cdot \pi \cdot \mathbf{Q}^{-T} \cdot \vec{r}_l) \equiv 1$$

by the construction via an orthonormal filter. \square .

A final remark concerns the smoothness of the built frames: similarly to the tight frames of RON, SHEN and GRÖCHENIG [98], the frames obtained by application of Theorem 4.13 can be made arbitrarily regular. We will show this fact now.

PROPOSITION* 4.15 *Consider a given tight frame $\Psi = \{ \varphi, \psi^1, \dots, \psi^m \}$ with frame bound $C_u = C_l = 1$, such that $\phi \in \mathcal{C}^\alpha$ for all $\phi \in \Psi$. Then all functions of the tight frame Ψ^j obtained by applying Theorem 4.13 to the generating filters of Ψ are in \mathcal{C}^β for some $\beta \geq j\alpha - 1 - \epsilon$ with $\epsilon > 0$ arbitrarily small.*

Proof. We only need to show the assertion for the scaling function φ^j , since all other $\psi^{i,j} \in \Psi^j$ can be written as finite linear combinations of the form

$$\psi^{i,j}(\vec{x}) = \sum_{\vec{l}} h_{i,\vec{l}} \cdot \varphi^j(\mathbf{Q}\vec{x} - \vec{l})$$

and are thus of the same regularity. We consider the FOURIER transform of $\varphi(\cdot)$ and assume that $\varphi(\cdot) \in \mathcal{C}^\alpha$. Since $\varphi(\cdot)$ is also compactly supported, we know from harmonic analysis that

$$|\hat{\varphi}(\vec{\omega})| \leq C_1 \cdot (1 + |\vec{\omega}|)^{-\alpha}, \quad \forall \vec{\omega} \in \mathbb{R}^n$$

must hold for some positive constant C_1 . On the other hand, the n -dimensional cascade algorithm (2.15) gives us

$$\hat{\varphi}(\vec{\omega}) = \hat{\varphi}(0) \cdot \prod_{l=1}^{\infty} H_0(\mathbf{Q}^{-lT} \cdot \vec{\omega}),$$

where $H_0(\vec{\omega})$ denotes the FOURIER transform of the related scaling filter. Now, we evaluate the modulus of $\hat{\varphi}^j$ and what we get is (with positive constants C_2, C_3)

$$\begin{aligned} \left| \hat{\varphi}^j(\vec{\omega}) \right| &= \left| \hat{\varphi}^j(0) \cdot \prod_{l=1}^{\infty} H_0(\mathbf{Q}^{-lT} \cdot \vec{\omega})^j \right| = \left| C_2 \cdot \hat{\varphi}(\vec{\omega})^j \right| \\ &\leq C_2 \cdot |\hat{\varphi}(\vec{\omega})|^j \leq C_3 \cdot (1 + |\vec{\omega}|)^{-j\alpha}, \quad \forall \vec{\omega} \in \mathbb{R}^n. \end{aligned}$$

Since φ^j is given by a finite number of convolutions of φ with itself, φ^j is also compactly supported. Thus, we have proven that all functions $\varphi^j \in \Psi^j$ must be in the HÖLDER class \mathcal{C}^β for some $\beta \geq j\alpha - 1 - \epsilon$. \square .

We additionally mention, that the estimate $j\alpha - 1 - \epsilon$ is only a lower bound for the HÖLDER exponent; the defacto value of β turns out to be much better in many cases. A — for bigger filters very time consuming — way to get more exact (sometimes even sharp) estimates for the smoothness value β is the application of Theorem 4.6.

Epilogue —

Which Class of Wavelets Should One Use Eventually?

To end with the chapter about wavelet design, we will briefly discuss, which of the presented wavelets — or more general — wavelet classes, should be taken into account when one is targeting at certain applications. Of course, we will focus ourselves to applications related to computer vision.

For tasks like image/signal compression or transmission, the most important requirement is to generate short codes in order to spend less storage space or short transmission time. This demand leads naturally to use orthogonal or biorthogonal wavelets, since these are minimal function families and thus also

create a minimal number of coefficients. Additionally, the (at least possibly) symmetric biorthogonal wavelets are better suited for lossy image compression, since the human visual system is more tolerant to symmetric errors than asymmetric ones; in fact, the most successful image coders (like SPIHT or the new JPEG2000 standard) work with symmetric biorthogonal wavelets.

But as soon as one does not only need to represent data in a possibly compact way but also wants to analyse the data or some of its substructures, a minimal number of coefficients is no longer a desirable property. Contrarily, it may even be a drawback, since redundancy of function systems leads to more robustness and higher error tolerance in numerical data analysis [164] [62]. This is also reflected by the fact, that for most low level vision tasks, redundant function systems like GAUSSIANS or GABOR frames are applied successfully. Consequently, we will follow this way and make mostly use of families of symmetric tight wavelet frames, which may be viewed as *discrete relatives* of GABOR functions. The intrinsic properties (vanishing moments, regularity, ...) of the applied frames will vary from case to case, depending on the concrete requirements. For simplicity, all the experiments in the third part of this thesis will be executed with the tight frame example given in the appendix about filter families.

Chapter Summary

After specifying some fundamental constraints for multidimensional wavelet design, we have introduced the concept of lowerdimensional subfilters and proven some results about how they are contained in higherdimensional scaling filters. From these results, we were able to derive optimal shapes for filters in multiple dimensions. Since it became clear that multidimensional wavelets are harder to find and more difficult to implement in practical applications, we argued why it still makes sense to use these latter ones. In this context, we have given a number of general guidelines that should be considered in the design procedure of scaling and wavelet filters. Because multidimensional filters usually leave some degrees of design freedom (one of their advantages), we have demonstrated a number of possibilities how these free coefficients might be utilized in order to optimize the resulting wavelets into the one or another direction. In the sequel, other design methods were discussed and compared to the methods developed within this work. This was followed by a brief presentation of the biorthogonal case and by the introduction of tight wavelet frames. Due to their redundancy, these latter function systems allowed the integration of several desirable properties which were exclusive in the (bi)orthogonal case. In order to apply this theory successfully, we have developed a new and very efficient filter design method for building tight wavelet frames and prove some related results regarding their symmetry and regularity properties. The chapter

was closed by a brief discussion about the right choice of wavelet family for certain kinds of applications.

II

WAVELET ANALYSIS IN HIGHER DIMENSIONS

Chapter 5

BASICS

Wir kommen nun zur Analysis, diesem kunstvollsten und am feinsten verzweigten Gebilde der mathematischen Wissenschaft.

—DAVID HILBERT

To apply multidimensional wavelet transforms successfully and effectively to applications of any kind, it is necessary to study its analytical properties first. Therefore, the goal of the actual chapter is the introduction of some fundamental wavelet based analytical tools. A very basic requirement to work with wavelets in real-world situations is to have a *suited* finite-dimensional approximation of the data to be represented. In the first section, the simplest form of such approximations are deduced from the operators that project functions into subspaces of L_2 . Section 2 introduces the MALLAT algorithm which calculates wavelet approximations in linear time depending on the data size. It is presented in a generalized form that also considers the multidimensional case using arbitrary integer scaling matrices. In the final section of this chapter, some improved wavelet approximations in more general function spaces are deduced. Most of the material presented within this chapter is standard (as e.g. in [140] [184] [74]) and only adapted here and there for the multidimensional case (mainly Proposition 5.3). The terminology and notation in this chapter is oriented to orthogonal wavelets just for simplicity of notation; the reader shall keep in mind that all the concepts are also applicable to tight wavelet frames.

5.1. Wavelet Approximation and the Projection Operators P_j

Recalling the definition of the discrete wavelet transform (Definition 2.4) and of a MRA (pages 17, 30), this transformation yields a representation of any

function $f \in L_2(\mathbb{R}^n)$ as a series of weighted functions in a nested sequence of subspaces V_j . For this reason, we define operators which represent these approximations of f on V_j .

DEFINITION 5.1 *Given a multidimensional MRA Ψ spanning a nested sequence*

$$\cdots \subset V_1 \subset V_0 \subset V_{-1} \subset \cdots \subset L_2(\mathbb{R}^n)$$

of subspaces. Then, for any $f \in L_2(\mathbb{R}^n)$ and $j \in \mathbb{Z}$ the projection operator

$$\mathbf{P}_j : L_2(\mathbb{R}^n) \longrightarrow V_j$$

is defined by

$$\mathbf{P}_j(f)(\vec{x}) = \sum_{\vec{k} \in \mathbb{Z}^n} \langle f, \varphi_{j,\vec{k}} \rangle \cdot \varphi_{j,\vec{k}}(\vec{x}). \quad (5.1)$$

$\mathbf{P}_j(f)$ is sometimes said to be the representation of f on the j -th scale and $\langle f, \varphi_{j,\vec{k}} \rangle$ is called the expansion coefficient of f with respect to $\varphi_{j,\vec{k}}$ (compare this to (2.3)).

By the construction of a MRA, we clearly have the limit cases

$$\lim_{j \rightarrow \infty} \|\mathbf{P}_j(f)\|_{L_2} = 0 \quad \text{and} \quad \lim_{j \rightarrow -\infty} \|\mathbf{P}_j(f) - f\|_{L_2} = 0. \quad (5.2)$$

As was shown earlier, for a MRA built from a \mathbf{Q} -adic dilation with $q = |\det \mathbf{Q}|$, the complement of the space V_j in V_{j-1} is spanned by $q-1$ spaces W_j^m with

$$V_{j-1} = V_j \oplus \bigcup_{m=1}^{q-1} W_j^m \quad \text{and} \quad V_j \perp W_j^m \quad \forall \quad m = 1, \dots, q-1.$$

Associated to the spaces W_j^m are the wavelets $\psi_j^m(\vec{x})$. This decomposition directly induces the definition of the complementary operators $\mathbf{U}_j^m(f)$ via

$$\mathbf{P}_{j-1}(f) = \mathbf{P}_j(f) + \sum_{m=1}^{q-1} \mathbf{U}_j^m(f). \quad (5.3)$$

Rewritten in terms of the transformation, the complementary projectors read

$$\mathbf{U}_j^m(f)(\vec{x}) = \sum_{\vec{k} \in \mathbb{Z}^n} \langle f, \psi_{j,\vec{k}}^m \rangle \cdot \psi_{j,\vec{k}}^m(\vec{x}). \quad (5.4)$$

This leads to the decomposition

$$\begin{aligned} f(\vec{x}) &= \mathbf{P}_j(f)(\vec{x}) + \sum_{k=-\infty}^j \sum_{m=1}^{q-1} \mathbf{U}_k^m(f)(\vec{x}) \\ &= \sum_{k \in \mathbb{Z}} \sum_{m=1}^{q-1} \mathbf{U}_k^m(f)(\vec{x}) \end{aligned} \quad (5.5)$$

for all $f \in L_2(\mathbb{R}^n)$ and all $j \in \mathbb{Z}$. The relations (5.2) and (5.5) imply that $\mathbf{P}_j(f)$ is indeed an approximation of f in the L_2 -sense, but they don't give any concrete information about the rate of convergence.

PROPOSITION 5.2 *Given a \mathbf{Q} -adic MRA built from a scaling function $\varphi(\vec{x})$ and $q - 1$ wavelets $\psi^m(\vec{x})$ such that*

$$\mu_{\vec{p}}^{\psi^m} = 0 \quad m = 1, \dots, q - 1$$

for all \vec{p} with $|\vec{p}| \leq \nu \in \mathbb{N}$. Then, for every compactly supported function f having at least ν continuous derivatives, the following estimate holds true

$$\|\mathbf{P}_j(f) - f\|_{L_2} \leq C \cdot |\det \mathbf{Q}|^{j \cdot \nu}. \quad (5.6)$$

Proof. The approximation via $\mathbf{P}_j(f)$ is very similar to finite element approximation and the error estimates can be proven in exactly the same way [85]. \square .

In practice, one usually samples a given signal $f \in L_2(\Omega)$ in order to work with the finite-dimensional discretized series $f_n \in l_2(\Omega \cap \mathbb{Z}^n)$. Therefore, it also might be useful to have a wavelet-based operator that approximates functions $f \in L_2$ by sampled function values rather than by transformation. For this reason, we define

$$\mathbf{S}_j(f)(\vec{x}) = |\det \mathbf{Q}|^{j/2} \cdot \sum_{\vec{k} \in \mathbb{Z}^n} \sum_{\vec{l} \in \mathbb{Z}^n} \lambda_{\vec{l}} \cdot f(\vec{l} + \mathbf{Q}^j \cdot \vec{k}) \cdot \varphi_{j,\vec{k}}(\vec{x}) \quad (5.7)$$

to be the *wavelet sampling operator* of step size \mathbf{Q}^j . Hereby, the weighting coefficients $\lambda_{\vec{l}}$ are obtained as the solution of the linear equation system

$$\sum_{\vec{l}} \lambda_{\vec{l}} \cdot \mu_{\vec{m}, \mathbf{Q}^{-j} \cdot \vec{l}}^{\varphi} = \delta_{\vec{m}}. \quad (5.8)$$

As in the case of the projectors $\mathbf{P}_j(f)$, the following convergence result is valid.

PROPOSITION* 5.3 *With the same hypotheses as in Proposition (5.2) and additionally assuming that $\varphi(\vec{x})$ is at least continuous and the dilation matrix \mathbf{Q} satisfies*

$$\mathbf{Q}^j = (\det \mathbf{Q})^{1/n} \cdot \mathbf{Id}, \quad (5.9)$$

we have the error estimation

$$\|\mathbf{S}_j(f) - f\|_{L_2} \leq C \cdot |\det \mathbf{Q}|^{j \cdot \nu}. \quad (5.10)$$

Proof. Postponed to Appendix 5.A. \square .

Remark 1. Similar results for the onedimensional case were introduced in [184] and [217]. In 1D, a criterion similar to (5.9) is not necessary. Even in higher dimensions, the restrictive condition (5.9) may be relaxed in the following sense: if there exists an increasing sequence $(s_p)_{p \in \mathbb{N}}$ with $\{s_p\} \subset j \cdot \mathbb{Z}$ such that

$$(\mathbf{S}_p)_{p \in \mathbb{N}} = \left((\det \mathbf{Q}^{-s_p/n}) \cdot \mathbf{Q}^{s_p} \right)_{p \in \mathbb{N}}$$

forms a CAUCHY sequence with respect to the usual matrix norm and tends to the limit

$$\lim_{p \rightarrow \infty} \mathbf{S}_p = \mathbf{Id},$$

then Proposition 5.3 holds also true (only the weights $\lambda_{\vec{\gamma}}$ have to be chosen differently). Since the proof of this relaxation would be very time consuming and involve additional techniques like diophantine approximation, it is not presented here. However, two contrary examples to this issue are discussed in Appendix 5.B. We shall finally remark that the condition (5.9) is not a too big obstruction in practice, since it is satisfied by all regular dilation matrices as well as by the quincunx matrices and its higherdimensional generalizations (see Proposition 3.7).

Remark 2. If one uses coiflets, i.e. $\mu_m^q = 0$ for $0 < |\vec{m}| \leq \nu$, it directly follows that (5.7) becomes the simple one-point quadrature formula

$$\mathbf{S}_j(f)(\vec{x}) = |\det \mathbf{Q}|^{j/2} \cdot \sum_{\vec{k} \in \mathbb{Z}^n} f(\mathbf{Q}^j \cdot \vec{k}) \cdot \varphi_{j,\vec{k}}(\vec{x}). \quad (5.11)$$

Although it is constructed by usage of coiflet scaling functions, the simpler *sampling scheme* (5.11) approximates $f(\vec{x})$ not only in the coiflet case, but for any arbitrary scaling function $\varphi(\vec{x})$ — only the convergence is slower in the non-coiflet case. This universality of (5.11) will play a role in the following section.

5.2. The Fast Matrix Dilation Wavelet Transform

The approximation properties of the sampling operator $\mathbf{S}_j(f)$ play a major role in practical considerations. Commonly, a sampled discrete data set is interpreted as a wavelet representation of the considered signal on some fixed scale j_0 (mostly taken to be $j_0 = 0$ for simplicity). This results in the idealized assumption

$$f(\vec{x}) \approx \mathbf{S}_0(f) = \sum_{\vec{k} \in \mathbb{Z}^n} c_{\vec{k}}^0 \cdot \varphi_{0,\vec{k}}(\vec{x}), \quad (5.12)$$

where the coefficients are the sampled function values, i.e. $c_{\vec{k}}^0 = f_{\vec{k}}$. We are now interested in finding an associated wavelet expansion of the form

$$f(\vec{x}) = \sum_{\vec{k} \in \mathbb{Z}^n} c_{\vec{k}}^i \cdot \varphi_{i,\vec{k}}(\vec{x}) + \sum_{p=1}^{|q|-1} \sum_{j=1}^i \sum_{\vec{k} \in \mathbb{Z}^n} d_{\vec{k}}^{j,p} \cdot \psi_{j,\vec{k}}^p(\vec{x}) \quad (5.13)$$

and therefore, it is necessary, to successively obtain the expansion coefficients $c_{\vec{k}}^j, d_{\vec{k}}^{j,p}$ for the higher scales. This is the discrete wavelet transform itself; we only need to evaluate the inner products in (5.1) and (5.4). This is done in the following way:

$$\begin{aligned} c_{\vec{l}}^{j+1} &= \sum_{\vec{k}} c_{\vec{k}}^j \cdot \int_{\mathbb{R}^n} \varphi_{j,\vec{k}}(\vec{x}) \cdot \varphi_{j-1,\vec{l}} d\vec{x} \\ &= |\det \mathbf{Q}|^{-1/2} \cdot \sum_{\vec{k}} \sum_{\vec{m}} c_{\vec{k}}^j \cdot h_{\vec{m}} \cdot \int_{\mathbb{R}^n} \varphi_{j,\vec{k}}(\vec{x}) \cdot \varphi_{j,\mathbf{Q} \cdot \vec{l} + \vec{m}} d\vec{x} \\ &= |\det \mathbf{Q}|^{-1/2} \cdot \sum_{\vec{k}} \sum_{\vec{m}} c_{\vec{k}}^j \cdot h_{\vec{m}} \cdot \delta_{\mathbf{Q} \cdot \vec{l} + \vec{m} + \vec{k}} \\ &= |\det \mathbf{Q}|^{-1/2} \cdot \sum_{\vec{k}} c_{\vec{k}}^j \cdot h_{\vec{k} - \mathbf{Q} \cdot \vec{l}}. \end{aligned} \quad (5.14)$$

Hereby, we used the scaling equation (3.7) and the pairwise orthonormality of the $\varphi_{\vec{k}}(\vec{x})$. The analogous relation for the wavelet coefficients $d_{\vec{k}}^{j+1,p}$ can be found in a similar manner, it is stated as

$$d_{\vec{k}}^{j+1,p} = |\det \mathbf{Q}|^{-1/2} \cdot \sum_{\vec{k}} d_{\vec{k}}^{j,p} \cdot g_{\vec{k} - \mathbf{Q} \cdot \vec{l}}^p \quad p = 1, \dots, |q| - 1. \quad (5.15)$$

This *wavelet decomposition* of the coefficients is very efficient, since it is done by a simple discrete convolution followed by a subsampling with the sampling matrix \mathbf{Q} . This procedure is known as *MALLAT's algorithm* in the wavelet literature. It makes the remarkable property of the discrete wavelet transformation evident, that one only needs to know the scaling and wavelet coefficients to work with the transform; no further knowledge about the wavelets themselves is necessary in any manner.

The *wavelet reconstruction* of a signal is the inverse process to what happens in (5.14) and (5.15). From the orthonormality of the filter, it directly follows that the reconstruction is done by the adjoint of the decomposition operators. Formally, this reads

$$c_{\vec{l}}^{j-1} = |\det \mathbf{Q}|^{-1/2} \cdot \left(\sum_{\vec{k}} c_{\vec{k}}^j \cdot h_{\mathbf{Q} \cdot \vec{k} - \vec{l}} + \sum_{p=1}^{|q|-1} \sum_{\vec{k}} d_{\vec{k}}^{j,p} \cdot g_{\mathbf{Q} \cdot \vec{k} - \vec{l}}^p \right). \quad (5.16)$$

A simple complexity analysis shows that the wavelet decomposition as well as the wavelet reconstruction algorithm both have a complexity of $O(N)$ operations, where N denotes the size of the data set. This makes it even faster than most standard integral transformations, even faster than the successful *fast FOURIER transform* (requiring $O(N \cdot \log N)$ operations). Additionally, the wavelet coefficients give a signal representation that is localized in time (space) and in frequency.

5.3. Generalized Wavelet Approximation

Although this may seem a paradox, all exact science is dominated by the idea of approximation.

—BERTRAND RUSSELL

This section is devoted to give a simple generalization of the convergence results presented in the first section of this chapter. It is the approximation in SOBOLEV spaces, which is of particular interest, since it plays a prominent role in solving ill-posed problems or systems of PDEs (see also Chapter 7).

Using the wavelet expansion coefficient characterization (B.11) of functions in SOBOLEV spaces, the approximation error of the projection operator $\mathbf{P}_j(f)$ can be evaluated to be

$$\|f - \mathbf{P}_j(f)\|_{W_p^s} = \left(\sum_{i \leq j} \sum_{m=1}^{|q|-1} \sum_{\vec{k}} |\det \mathbf{Q}|^{i \cdot (p-2) + i \cdot s \cdot p} \cdot |d_{\vec{k}}^{i,m}|^p \right)^{1/p}. \quad (5.17)$$

Under application of the HÖLDER inequality, one immediately gets the following approximation order estimation in terms of the SOBOLEV norm.

COROLLARY 5.4 *Let $f \in W_p^s$ and a MRA with $\varphi \in W_p^t$ such that $0 \leq s < t$ be given. Then the approximation error of $\mathbf{P}_j(f)$ is bounded by*

$$\|f - \mathbf{P}_j(f)\|_{W_p^u} = C \cdot |\det \mathbf{Q}|^{j \cdot (s-u)} \cdot \|f\|_{W_p^s} \quad (5.18)$$

for all $-t < u < t$ with $u \leq s$.

Equations like (5.18) are sometimes called JACKSON-type inequalities due to their similarity to error estimations in polynomial approximation established by D. JACKSON. Unfortunately, these error estimates do not give theoretical optimal approximation orders for smooth functions, but at least for the spaces W_2^s we can derive the following result.

COROLLARY* 5.5 *Under the conditions of Proposition 5.2, we have the estimate*

$$\|\mathbf{P}_j(f) - f\|_{W_2^s} \leq C \cdot |\det \mathbf{Q}|^{j \cdot (v-s)} \quad (5.19)$$

for integer values s . If the additional conditions of Proposition 5.3 are also met, the same estimate holds true for the operator $\mathbf{S}_j(f)$.

Proof. This result will become clear as an immediate consequence of Corollary 6.8. \square .

For regularization purposes, we will also use *best approximations* of f with respect to BESOV norms. However, this kind of approximations will not directly depend on the projectors $\mathbf{P}_j(f)$; they will be introduced in Chapter 8.

Chapter Summary

The wavelet projection and sampling operators were introduced. These build the basis for any sensible approximation done prior to a wavelet transform. Some important approximation results were given (and one will be proven in an upcoming appendix). Furthermore, the closely related fast matrix dilation wavelet transform was also presented. In a second appendix, a well-behaving and a pathological case for the approximation order of the sampling operator under two certain non-regular dilation matrices will be compared.

APPENDIX 5.A: Proof of Proposition 5.3

By expanding f into a TAYLOR series in the sampling points, that is, developing around $f(\vec{l} + \mathbf{Q}^j \cdot \vec{k})$, one obtains

$$\begin{aligned} \mathbf{S}_j(f)(\vec{x}) &= |\det \mathbf{Q}|^{j/2} \cdot \sum_{\vec{k} \in \mathbb{Z}^n} \sum_{\vec{l} \in \mathbb{Z}^n, |\vec{l}| \leq \nu} \lambda_{\vec{l}} \cdot f(\vec{l} + \mathbf{Q}^j \cdot \vec{k}) \cdot \varphi_{j,\vec{k}}(\vec{x}) \\ &= |\det \mathbf{Q}|^{j/2} \cdot \sum_{\vec{k} \in \mathbb{Z}^n} \sum_{\vec{l} \in \mathbb{Z}^n, |\vec{l}| \leq \nu} \lambda_{\vec{l}} \cdot \left(\sum_{\vec{m} \in \mathbb{N}^n, |\vec{m}| < \nu} \frac{1}{\vec{m}!} \cdot \mathbf{D}_{\vec{m}} f(\vec{x}) \cdot (\vec{l} + \mathbf{Q}^j \cdot \vec{k} - \vec{x})^{\vec{m}} \right. \\ &\quad \left. + \sum_{\vec{m} \in \mathbb{N}^n, |\vec{m}| = \nu} \frac{1}{\vec{m}!} \cdot \mathbf{D}_{\vec{m}} f(\vec{x}) \cdot (\vec{l} + \mathbf{Q}^j \cdot \vec{k} - \vec{x})^{\vec{m}} \right) \cdot \varphi_{j,\vec{k}}(\vec{x}) \end{aligned}$$

with some well chosen $\vec{x} \in \Omega$. We shall assume for the moment that

$$(*) = \sum_{\vec{k}} \sum_{\vec{l}} \lambda_{\vec{l}} \cdot (\vec{l} + \mathbf{Q}^j \cdot \vec{k} - \vec{x})^{\vec{m}} \cdot \varphi_{j,\vec{k}}(\vec{x}) \equiv C \cdot \delta_{\vec{l}} \quad \text{for } |\vec{l}| \leq \nu \quad (5.A.1)$$

with some constant C holds true. Then the representation of $\mathbf{S}_j(f)$ simplifies to

$$\mathbf{S}_j(f)(\vec{x}) = f(\vec{x}) + |\det \mathbf{Q}|^{j/2} \cdot \sum_{\vec{k}} \sum_{\vec{l}} \sum_{\vec{m}} \frac{\lambda_{\vec{l}}}{\vec{m}!} \cdot \mathbf{D}_{\vec{m}} f(\vec{x}) \cdot (\vec{l} + \mathbf{Q}^j \cdot \vec{k} - \vec{x})^{\vec{m}} \cdot \varphi_{j,\vec{k}}(\vec{x}).$$

From this, we can directly deduce the error description as

$$\begin{aligned}
\|S_j(f) - f\|_{L_2} &= |\det \mathbf{Q}|^{j/2} \cdot \left\| \sum_{\vec{k}} \sum_{\vec{l}} \sum_{\vec{m}} \frac{\lambda_{\vec{l}}}{\vec{m}!} \cdot \mathbf{D}_{\vec{m}} f(\xi \vec{k}) \cdot (\vec{l} + \mathbf{Q}^j \cdot \vec{k} - \vec{x})^{\vec{m}} \cdot \varphi_{j, \vec{k}}(\vec{x}) \right\|_{L_2} \\
&= |\det \mathbf{Q}|^{j/2} \cdot \left\| \sum_{\vec{k}} \sum_{\vec{l}} \sum_{\vec{m}} \frac{\lambda_{\vec{l}}}{\vec{m}!} \cdot \mathbf{D}_{\vec{m}} f(\xi \vec{k}) \cdot (\vec{l} + \mathbf{Q}^j \cdot \vec{x})^{\vec{m}} \right\|_{L_2} \\
&\leq |\det \mathbf{Q}|^{j/2 + j \cdot \nu} \cdot \left((2 \cdot s)^n \cdot (\mathbf{Q}^{-j} + 1) \cdot \int_{\Omega} \left(\sum_{\vec{l}} \sum_{\vec{m}} \frac{\lambda_{\vec{l}}}{\vec{m}!} \cdot \max_{\xi} |\mathbf{D}_{\vec{m}} f(\xi)| \cdot (\mathbf{Q}^{-j} \cdot \vec{l} + \vec{x})^{\vec{m}} \right)^2 d\vec{x} \right)^{1/2}
\end{aligned}$$

This last estimation is valid, since f is compactly supported and we may thus assume that $\text{supp}(f) \subset [-s, s]^n$ and $\text{supp}(\varphi) \subset [-s, s]^n$ for some $s > 0$. Furthermore, we note that the double sum in the integral is finite and the polynomial $(\mathbf{Q}^{-j} \cdot \vec{l} + \vec{x})^{\vec{m}}$ is uniformly bounded on Ω . This finally results in

$$\|S_j(f) - f\|_{L_2} \leq C \cdot |\det \mathbf{Q}|^{j \cdot \nu},$$

with a constant C that only depends on f and Ω . We are completely done with the proof, if we can verify the assumption (5.A.1). For this reason, consider for any given \vec{m} the moments

$$\begin{aligned}
\mu_{\vec{m}}^{\varphi} &= \sum_{\vec{k}} h_{\vec{k}} \cdot \vec{k}^{\vec{m}} \\
&= |\det \mathbf{Q}|^{-1/2} \cdot i^{|\vec{m}|} \cdot \mathbf{D}_{\vec{m}} H(\vec{\omega}) \Big|_{\vec{\omega} = 2 \cdot \pi \cdot \vec{k}} \quad \vec{k} \in \mathbb{Z}^n.
\end{aligned}$$

The second equality follows from (3.9). Application of (3.8) leads to

$$i^{|\vec{m}|} \cdot \mathbf{D}_{\vec{m}} \hat{\varphi}(\omega) \Big|_{\vec{\omega} = 2 \cdot \pi \cdot \vec{k}} = \delta_{\vec{k}} \cdot \mu_{\vec{m}}^{\varphi},$$

which yields — under usage of the POISSON summation formula — the identity

$$\mu_{\vec{m}}^{\varphi} \equiv \sum_{\vec{k}} (\vec{x} - \vec{k})^{\vec{m}} \cdot \varphi(\vec{x} - \vec{k}). \quad (5.A.2)$$

By a simple translation, a similar identity assertion to (5.A.2) can be given for shifted moments. Application of the condition (5.9) and inserting (5.A.2) into (5.A.1) finally gives

$$\begin{aligned}
(*) &= |\det \mathbf{Q}^{-j/2}| \cdot \sum_{\vec{k}} \sum_{\vec{l}} \lambda_{\vec{l}} \cdot (\vec{l} + \mathbf{Q}^j \cdot \vec{k} - \vec{x})^{\vec{m}} \cdot \varphi(\mathbf{Q}^{-j} \cdot \vec{x} - \vec{k}) \\
&= \pm |\det \mathbf{Q}^{-3j/2}| \cdot (-1)^{|\vec{m}|} \cdot \sum_{\vec{l}} \lambda_{\vec{l}} \cdot \sum_{\vec{k}} (\mathbf{Q}^{-j} \cdot \vec{x} - \vec{k} - \mathbf{Q}^{-j} \cdot \vec{l})^{\vec{m}} \cdot \varphi(\mathbf{Q}^{-j} \cdot \vec{x} - \vec{k}) \\
&= \pm |\det \mathbf{Q}^{-3j/2}| \cdot (-1)^{|\vec{m}|} \cdot \sum_{\vec{l}} \lambda_{\vec{l}} \cdot \mu_{\vec{m}, \mathbf{Q}^{-j} \cdot \vec{l}}^{\varphi} \\
&= |\det \mathbf{Q}^{-3j/2}| \cdot \delta_{\vec{m}}
\end{aligned}$$

by the presumption (5.8), which finishes the proof. \square .

APPENDIX 5.B: Two examples

We will give one example and one counterexample for the relaxed condition remarked following Proposition 5.3: first consider the dyadic dilation

$$\mathbf{Q} = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix}.$$

Observe, that the eigenvalues of \mathbf{Q} are given by

$$\lambda_{1,2} = \frac{1}{2} \pm \frac{i}{2} \cdot \sqrt{7} = \sqrt{2} \cdot e^{\pm i \cdot \arctan \sqrt{7}}.$$

Since the exponent is an irrational number, the eigenvalues $\lambda_{1,2}^n$ of \mathbf{Q}^n lie dense on the circle of radius $\sqrt{2}$. By a theorem of SIEGEL about diophantine approximation [208], there always exists a subsequence $(l_k)_{k \in \mathbb{N}}$ of $(\lambda_{1,2}^n)_{n \in \mathbb{N}}$ such that l_k converges to $\sqrt{2}$ and since $\lambda_1 = \overline{\lambda_2}$ the associated normed matrix powers \mathbf{S}_{l_k} converge to the identity operator. For example, taking $j = 5$ one obtains a sequence

$$k = 5, 130, 135, 265, 400, 665, 1065, 2795, 31810, 34605, 101020, 135625, 2406645, \dots,$$

for which $\arg(\lambda_{1,2})$ decreases and $\lambda_{1,2}^k$ converges against $\sqrt{2}$. For some values of k the matrix powers \mathbf{Q}^k are

$$\begin{aligned} \mathbf{Q}^5 &\approx 2^{5/2} \cdot \begin{bmatrix} 0.8838 & 0.1767 \\ -0.3535 & 1.0607 \end{bmatrix}, \quad \mathbf{Q}^{400} \approx 2^{200} \cdot \begin{bmatrix} 0.9867 & 0.0254 \\ -0.0508 & 1.0121 \end{bmatrix}, \\ \mathbf{Q}^{2795} &\approx 2^{k_8/2} \cdot \begin{bmatrix} 1.0003 & -0.0007 \\ 0.0014 & 0.9996 \end{bmatrix}, \quad \mathbf{Q}^{2406645} \approx 2^{k_{13}/2} \cdot \begin{bmatrix} 0.9999962 & 0.0000077 \\ -0.0000154 & 1.0000038 \end{bmatrix}. \end{aligned}$$

For the second example, which does not satisfy the relaxed convergence condition, take the tetra-adic dilation matrix

$$\mathbf{Q} = \begin{bmatrix} 2 & q \\ 0 & 2 \end{bmatrix}$$

with some $q \in \mathbb{Z}$. By a simple calculation, one evaluates the k -th matrix power of \mathbf{Q} to be

$$\mathbf{Q}^k = 2^k \cdot \begin{bmatrix} 1 & \frac{q \cdot k}{2} \\ 0 & 1 \end{bmatrix}$$

and it easily follows that this can obviously neither converge to a multiple of the identity operator nor does any subsequence so as long as $q \neq 0$.

Chapter 6

OPERATORS IN BASES OF NONSEPARABLE WAVELETS

... there are large classes of operators for which wavelets are nearly their eigenvectors.

—PHILIPPE TCHAMITCHIAN [219]

In this chapter, we will introduce some material concerning the application of wavelet-based operators. The focus will be set on such operators that will be useful for our computer vision purposes, although similar results to those presented here are also valid for a much broader class of operators. We consider an operator \mathbf{T} as an integral transform

$$\mathbf{T}(f)(\vec{x}) = \int_{\mathbb{R}^n} k(\vec{x}, \vec{y}) \cdot f(\vec{y}) \, d\vec{y}, \quad (6.1)$$

with the *transformation kernel* $k(\vec{x}, \vec{y})$. The quotation that introduces this chapter already suggested that wavelet-based operators have the nice property, that their eigenvectors are *close* to those wavelets the operator is built from — a very useful feature that is explained by the good localization wavelets have in time and frequency. In terms of the transformation kernel k , this means that $k(\vec{x}, \vec{y})$ is strong diagonally dominant, which makes (among other things) the efficient calculation of inverse operators possible.

In the first section, we will discuss the properties of differential operators in the wavelet domain and introduce *connection coefficients*, which represent the discrete version of the differential operator kernel k . Together with the approximation properties of wavelets derived in the previous chapter, this will yield approximative characterizations for functional derivatives and also allows us to relate this to classical finite difference schemes. Section 6.2 deals with affine transformations and a short sketch about RIESZ transforms in wavelet

coordinates. Mainly, we will obtain a technique for efficient image warping based on an efficient calculation of translational operators in the wavelet domain. In the third section, we will introduce the two different ways how to represent operators in wavelet bases, namely the so-called *standard* and *non-standard* representation. This chapter closes with some brief remarks about more general operator classes — in particular CALDERÓN-ZYGMUND operators — and some related work is quoted.

6.1. Differential Operators and Connection Coefficients

In the previous chapter, we have shown how to use wavelet series for the representation of functions in certain spaces. Such a representation leads to an infinite series on the theoretical level or to a finite-dimensional approximation of arbitrary precision in a more practical context. As a natural question, the requirement for approximations of functional derivatives also arises in this context. Such derivatives are of essential value for doing calculus or solving problems of integro-differential type within the wavelet framework.

6.1.1. Definition and Fundamental Properties

Following this plan, we will now investigate the representation of differential operators in the wavelet domain. In particular, we consider operators of the type

$$\mathbf{D}_{\vec{m}}(f)(\vec{x}) = \frac{\partial^{|\vec{m}|}}{\partial \vec{x}^{\vec{m}}} f(\vec{x}), \quad (6.2)$$

and will develop a representation of $\mathbf{D}_{\vec{m}}(f)$ as a wavelet series

$$\mathbf{D}_{\vec{m}}(f)(\vec{x}) \approx \sum_{\vec{k} \in \mathbb{Z}^n} u_{\vec{k}}^i \cdot \varphi_{i,\vec{k}}(\vec{x}) + \sum_{p=1}^{|\vec{q}|-1} \sum_{j=1}^i \sum_{\vec{k} \in \mathbb{Z}^n} v_{\vec{k}}^{j,p} \cdot \psi_{j,\vec{k}}^p(\vec{x}) \quad (6.3)$$

similar to (5.13). Hereby, the wavelet expansion coefficients are given by

$$\begin{aligned} u_{\vec{k}}^i &= \int_{\mathbb{R}^n} \mathbf{D}_{\vec{m}}(f)(\vec{x}) \cdot \varphi_{i,\vec{k}}(\vec{x}) \, d\vec{x} \quad \text{and} \\ v_{\vec{k}}^{j,p} &= \int_{\mathbb{R}^n} \mathbf{D}_{\vec{m}}(f)(\vec{x}) \cdot \psi_{j,\vec{k}}^p(\vec{x}) \, d\vec{x}. \end{aligned} \quad (6.4)$$

Assuming $f(\vec{x})$ has in fact a representation (5.13), we may insert it into the defining relation (6.4) yielding

$$\begin{aligned} u_{\vec{k}}^i &= \sum_{\vec{l} \in \mathbb{Z}^n} c_{\vec{l}}^i \cdot \int_{\mathbb{R}^n} \mathbf{D}_{\vec{m}}(\varphi_{i,\vec{l}})(\vec{x}) \cdot \varphi_{i,\vec{k}}(\vec{x}) d\vec{x} + \\ &\quad \sum_{p=1}^{|q|-1} \sum_{j=1}^i \sum_{\vec{l} \in \mathbb{Z}^n} d_{\vec{l}}^{j,p} \cdot \int_{\mathbb{R}^n} \mathbf{D}_{\vec{m}}(\psi_{j,\vec{l}}^p)(\vec{x}) \cdot \varphi_{i,\vec{k}}(\vec{x}) d\vec{x} \end{aligned} \quad (6.5)$$

and an analogous expression for the coefficients $v_{\vec{k}}^{j,p}$. That means, we can compute the derivatives of f if we are able to compute the inner products between the scaling functions/wavelets and their partial derivatives. This motivates to define

DEFINITION* 6.1 *Let $\{\varphi, \psi^1, \dots, \psi^{|q|-1}\}$ form a MRA with a integer dilation matrix \mathbf{Q} with $q = \det \mathbf{Q}$. Then, the inner products*

$$\begin{aligned} \Gamma_{\varphi, \vec{m}}^{\varphi}(i, \vec{k}, j, \vec{l}, \nu) &= \langle \mathbf{D}_{\mathbf{Q}^{-\nu} \cdot \vec{m}} \varphi_{i,\vec{k}}, \varphi_{j,\vec{l}} \rangle, \\ \Gamma_{\varphi, \vec{m}}^{\psi^p}(i, \vec{k}, j, \vec{l}, \nu) &= \langle \mathbf{D}_{\mathbf{Q}^{-\nu} \cdot \vec{m}} \varphi_{i,\vec{k}}, \psi_{j,\vec{l}}^p \rangle, \\ \Gamma_{\psi^p, \vec{m}}^{\varphi}(i, \vec{k}, j, \vec{l}, \nu) &= \langle \mathbf{D}_{\mathbf{Q}^{-\nu} \cdot \vec{m}} \psi_{i,\vec{k}}^p, \varphi_{j,\vec{l}} \rangle, \\ \Gamma_{\psi^p, \vec{m}}^{\psi^r}(i, \vec{k}, j, \vec{l}, \nu) &= \langle \mathbf{D}_{\mathbf{Q}^{-\nu} \cdot \vec{m}} \psi_{i,\vec{k}}^p, \psi_{j,\vec{l}}^r \rangle \end{aligned}$$

are called connection coefficients of order \vec{m} and exponent ν for wavelet differentiation.

Obviously, the connection coefficients are the discrete version of the transformation kernel of the differential operator $\mathbf{D}_{\mathbf{Q}^{-\nu} \cdot \vec{m}}$. This idea originally goes back to BEYLKIN [21], while the term connection coefficient in conjunction with wavelets was formally introduced in [132]. The multidimensional, matrix dilation generalization of this concept is due to the author and was first presented in [170].

In order to work with the connection coefficients, we have to calculate them first. One possibility is to evaluate the integrals numerically by suited quadrature formulae, but such a proceeding might cause numerical inaccuracies as well. Therefore, we will develop a different way using the special properties of wavelets that gives the exact values for the connection coefficients by solving a linear system of equations. This requires some lemmata summarizing the elementary properties of connection coefficients.

LEMMA* 6.2 With the preliminaries of Definition 6.1, the following relations are valid

$$\begin{aligned}\Gamma_{\varphi, \vec{m}}^{\varphi}(\vec{i}, \vec{k}, j, \vec{l}, \nu) &= (-1)^{|\vec{m}|} \cdot \Gamma_{\varphi, \vec{m}}^{\varphi}(j, \vec{l}, \vec{i}, \vec{k}, \nu), \\ \Gamma_{\varphi, \vec{m}}^{\psi^p}(\vec{i}, \vec{k}, j, \vec{l}, \nu) &= (-1)^{|\vec{m}|} \cdot \Gamma_{\psi^p, \vec{m}}^{\varphi}(j, \vec{l}, \vec{i}, \vec{k}, \nu), \\ \Gamma_{\psi^p, \vec{m}}^{\psi^r}(\vec{i}, \vec{k}, j, \vec{l}, \nu) &= (-1)^{|\vec{m}|} \cdot \Gamma_{\psi^p, \vec{m}}^{\psi^r}(j, \vec{l}, \vec{i}, \vec{k}, \nu).\end{aligned}\quad (6.6)$$

Proof. Since all scaling functions and wavelets tend to zero as $|\vec{x}|$ tends to infinity by definition, the multidimensional integration by parts rule and the transformation formula immediately give the result. \square .

LEMMA* 6.3 With the same assumptions as before and any decompositions $\vec{k} = \vec{k}_1 + \vec{k}_2$ and $\vec{l} = \vec{l}_1 + \vec{l}_2$, the following holds true.

$$\Gamma_{\varphi, \vec{m}}^{\varphi}(\vec{i}, \vec{k}, j, \vec{l}, \nu) = \Gamma_{\varphi, \vec{m}}^{\varphi}(\vec{i}, \vec{k}_1 - \mathbf{Q}^{j-\vec{i}} \cdot \vec{l}_2, j, \vec{l}_1 - \mathbf{Q}^{i-j} \cdot \vec{k}_2, \nu), \quad (6.7)$$

$$\Gamma_{\varphi, \vec{m}}^{\varphi}(\vec{i}, \vec{k}, j, \vec{l}, \nu + 1) = |\det \mathbf{Q}|^{-1} \cdot \Gamma_{\varphi, \vec{m}}^{\varphi}(\vec{i} - 1, \vec{k}, j - 1, \vec{l}, \nu). \quad (6.8)$$

Analogous relations are also valid for $\Gamma_{\varphi, \vec{m}}^{\psi}$, Γ_{ψ}^{φ} and Γ_{ψ}^{ψ} .

Proof. Application of the transformation formula with the substitution

$$\vec{x} \longrightarrow \vec{x} + \mathbf{Q}^j \cdot \vec{l}_2 + \mathbf{Q}^i \cdot \vec{k}_2$$

directly gives the relation (6.7) and similarly, the substitution $\vec{x} \longrightarrow \mathbf{Q}^{-1} \cdot \vec{x}$ gives (6.8). \square .

LEMMA* 6.4 Let $h_{\vec{k}}$ and $g_{\vec{k}}^p$ denote the scaling and wavelet coefficients. Then we have

$$\begin{aligned}\Gamma_{\varphi, \vec{m}}^{\varphi}(\vec{i} + 1, \vec{k}, j, \vec{l}, \nu) &= |\det \mathbf{Q}|^{1/2} \cdot \sum_{\vec{q}} h_{\vec{q}} \cdot \Gamma_{\varphi, \vec{m}}^{\varphi}(\vec{i}, \mathbf{Q} \cdot \vec{k} + \vec{q}, j, \vec{l}, \nu), \\ \Gamma_{\varphi, \vec{m}}^{\psi^p}(\vec{i} + 1, \vec{k}, j, \vec{l}, \nu) &= |\det \mathbf{Q}|^{1/2} \cdot \sum_{\vec{q}} h_{\vec{q}} \cdot \Gamma_{\varphi, \vec{m}}^{\psi^p}(\vec{i}, \mathbf{Q} \cdot \vec{k} + \vec{q}, j, \vec{l}, \nu), \\ \Gamma_{\psi^p, \vec{m}}^{\varphi}(\vec{i} + 1, \vec{k}, j, \vec{l}, \nu) &= |\det \mathbf{Q}|^{1/2} \cdot \sum_{\vec{q}} g_{\vec{q}}^p \cdot \Gamma_{\varphi, \vec{m}}^{\varphi}(\vec{i}, \mathbf{Q} \cdot \vec{k} + \vec{q}, j, \vec{l}, \nu), \\ \Gamma_{\psi^p, \vec{m}}^{\psi^r}(\vec{i} + 1, \vec{k}, j, \vec{l}, \nu) &= |\det \mathbf{Q}|^{1/2} \cdot \sum_{\vec{q}} g_{\vec{q}}^r \cdot \Gamma_{\varphi, \vec{m}}^{\psi^p}(\vec{i}, \mathbf{Q} \cdot \vec{k} + \vec{q}, j, \vec{l}, \nu).\end{aligned}\quad (6.9)$$

Proof. The results in (6.8) follow directly from the application of the scaling equation \square .

Combining these three lemmata, it is possible to calculate all connection coefficients once the *fundamental connection coefficients*

$$\Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \vec{l}, \nu)$$

are known, and it remains the task to evaluate them.

6.1.2. Calculation of Fundamental Connection Coefficients

In order to obtain a unique linear equation system, from which the connection coefficients are available, two additional propositions are necessary.

PROPOSITION* 6.5 *With the same notation and assumptions as before, we have the following identity*

$$\Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \vec{k}, \nu) = \sum_{\vec{l}} \sum_{\vec{q}} h_{\vec{l}} \cdot h_{\vec{q}} \cdot \Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \mathbf{Q} \cdot \vec{k} - \vec{l} + \vec{q}, \nu+1). \quad (6.10)$$

Proof. To show this result essentially bases again on the scaling equation and some integral calculus. We have

$$\begin{aligned} \Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \vec{k}, \nu) &= \int_{\mathbb{R}^n} \mathbf{D}_{\mathbf{Q}^{-\nu} \cdot \vec{m}} \varphi(\vec{x}) \cdot \varphi(\vec{x} - \vec{k}) d\vec{x} \\ &= |\det \mathbf{Q}| \cdot \sum_{\vec{l}} \sum_{\vec{q}} h_{\vec{l}} \cdot h_{\vec{q}} \cdot \int_{\mathbb{R}^n} \mathbf{D}_{\mathbf{Q}^{-\nu} \cdot \vec{m}} \varphi(\mathbf{Q} \cdot \vec{x} - \vec{l}) \cdot \\ &\quad \varphi(\mathbf{Q} \cdot \vec{x} - \mathbf{Q} \cdot \vec{k} - \vec{q}) d\vec{x} \\ &= \sum_{\vec{l}} \sum_{\vec{q}} h_{\vec{l}} \cdot h_{\vec{q}} \cdot \int_{\mathbb{R}^n} \mathbf{D}_{\mathbf{Q}^{-\nu-1} \cdot \vec{m}} \varphi(\vec{x}) \cdot \\ &\quad \varphi(\vec{x} - \mathbf{Q} \cdot \vec{k} + \vec{l} - \vec{q}) d\vec{x}, \end{aligned}$$

which is exactly the desired relation. \square .

PROPOSITION* 6.6 *If the scaling filter $(h_{\vec{k}})_{\vec{k} \in \mathbb{Z}^n}$ satisfies the sum rules of order $|\vec{m}|$ (see also (4.9)), then for all \vec{q} with $|\vec{q}| \leq |\vec{m}|$, we have*

$$\sum_{\vec{k}} \vec{k}^{\vec{q}} \cdot \Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \vec{k}, \nu) = (-1)^{|\vec{q}|} \cdot \vec{q}! \cdot \delta_{\vec{q}, \vec{m}}. \quad (6.11)$$

Proof. Since the scaling filter satisfies the according sum rules, the related wavelets have \vec{q} vanishing moments by Proposition 4.5. This means $\vec{x}^{\vec{q}} \in V_0$

and leads to a wavelet expansion

$$\begin{aligned}
\vec{x}^{\vec{q}} &= \sum_{\vec{k}} \int_{\mathbb{R}^n} \vec{y}^{\vec{q}} \cdot \varphi(\vec{y} - \vec{k}) d\vec{y} \cdot \varphi(\vec{x} - \vec{k}) \\
&= \sum_{\vec{k}} \int_{\mathbb{R}^n} (\vec{y} + \vec{k})^{\vec{q}} \cdot \varphi(\vec{y}) d\vec{y} \cdot \varphi(\vec{x} - \vec{k}) \\
&= \sum_{\vec{k}} \sum_{\vec{l}} \int_{\mathbb{R}^n} \binom{\vec{q}}{\vec{l}} \cdot \vec{k}^{\vec{q}-\vec{l}} \cdot \vec{y}^{\vec{l}} \cdot \varphi(\vec{y}) d\vec{y} \cdot \varphi(\vec{x} - \vec{k}) \\
&= \sum_{\vec{k}} \sum_{\vec{l}} \binom{\vec{q}}{\vec{l}} \cdot \vec{k}^{\vec{q}-\vec{l}} \cdot \mu_{\vec{l}}^{\varphi} \cdot \varphi(\vec{x} - \vec{k}).
\end{aligned}$$

Next, we apply the operator $\mathbf{D}_{\vec{m}}$ to both sides of this equation, multiply by $\varphi(\vec{x})$ and integrate over \mathbb{R}^n obtaining

$$\begin{aligned}
\vec{q}! \cdot \delta_{\vec{q}, \vec{m}} &= \int_{\mathbb{R}^n} \vec{q}! \cdot \delta_{\vec{q}, \vec{m}} \cdot \varphi(\vec{x}) \\
&= \int_{\mathbb{R}^n} \sum_{\vec{k}} \sum_{\vec{l}} \binom{\vec{q}}{\vec{l}} \cdot \vec{k}^{\vec{q}-\vec{l}} \cdot \mu_{\vec{l}}^{\varphi} \cdot \mathbf{D}_{\vec{m}} \varphi(\vec{x} - \vec{k}) \cdot \varphi(\vec{x}) d\vec{x} \\
&= \sum_{\vec{k}} \sum_{\vec{l}} \binom{\vec{q}}{\vec{l}} \cdot \vec{k}^{\vec{q}-\vec{l}} \cdot \mu_{\vec{l}}^{\varphi} \cdot \Gamma_{\varphi}^{\varphi}(0, \vec{k}, 0, \vec{0}, 0) \\
&= (-1)^{|\vec{q}|} \cdot \sum_{\vec{l}} \binom{\vec{q}}{\vec{l}} \cdot \mu_{\vec{l}}^{\varphi} \cdot \sum_{\vec{k}} \vec{k}^{\vec{q}-\vec{l}} \cdot \Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \vec{k}, 0)
\end{aligned}$$

By Lemma 6.2, the inner sum vanishes for all \vec{l} with $\vec{l} \neq \vec{0}$ and for $\vec{l} = \vec{0}$ it vanishes at least as long as $\vec{q} \neq \vec{m}$. Since $\mu_{\vec{0}}^{\varphi} = 1$, we have shown the claim for $\nu = 0$. For $\nu \neq 0$, the proof is completely analogous, one only has to replace $\vec{x}^{\vec{q}}$ by $(\mathbf{Q}^{\nu} \cdot \vec{x})^{\vec{q}}$. \square

Now, we can conclude our investigations about the properties of connection coefficients.

THEOREM* 6.7 *If a given scaling filter is finite and the dilation matrix of a MRA satisfies*

$$\mathbf{Q}^{\rho} \in \text{Diag}(n \times n)$$

for some $\rho \in \mathbb{N}$, then all fundamental connection coefficients are exactly calculable by a finite set of linear equations.

Proof. Proposition 6.5 makes it possible to calculate $\Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \vec{k}, \nu + 1)$ out of a set of linear equations in $\Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \vec{k}, \nu)$, then $\Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \vec{k}, \nu + 2)$

out of $\Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \vec{k}, \nu + 1)$ and so on. To *close* this chain of equation systems, we need to find a set of linear relations to calculate $\Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \vec{k}, 0)$ out of $\Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \vec{k}, \sigma)$ for some suited $\sigma \in \mathbb{N}$. Suppose now, that

$$\mathbf{Q}^{\rho} = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

and $\vec{\lambda} = [\lambda_1, \dots, \lambda_n]^T$, then it is directly verified that

$$\mathbf{D}_{\mathbf{Q}^{-\rho}, \vec{m}} = \vec{\lambda}^{\vec{m}} \cdot \mathbf{D}_{\vec{m}}.$$

This can be directly plugged into (6.10) yielding

$$\begin{aligned} \Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \vec{k}, \rho - 1) &= \sum_{\vec{l}} \sum_{\vec{q}} h_{\vec{l}} \cdot h_{\vec{q}} \cdot \Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \mathbf{Q} \cdot \vec{k} - \vec{l} + \vec{q}, \rho) \\ &= \vec{\lambda}^{\vec{m}} \cdot \sum_{\vec{l}} \sum_{\vec{q}} h_{\vec{l}} \cdot h_{\vec{q}} \cdot \Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{0}, 0, \mathbf{Q} \cdot \vec{k} - \vec{l} + \vec{q}, 0), \end{aligned}$$

which gives the desired relationship. Since the equations given by (6.10) lead to one degree of underdetermination for each differential exponent ν , the additional identities (6.11) for $\nu = 0, 1, \dots, \rho - 1$ make the system unique. \square .

Remark 1. Connection coefficients are a very useful and effective tool, since they can be computed offline (which may take quite some time) and stored into lookup-tables afterwards. This makes them directly and very fast accessible for all times once they are known.

Remark 2. In a very similar manner, connection coefficients can be generalized in several ways. One such example are *g-factor connection coefficients* defined by

$$\Gamma(j_1, \vec{k}_1, \vec{m}_1, \dots, j_g, \vec{k}_g, \vec{m}_g) = \int_{\mathbb{R}^n} \prod_{i=1}^g \mathbf{D}_{\vec{m}_i} \varphi_{j_i, \vec{k}_i}(\vec{x}) d\vec{x}.$$

Another useful extension are *polynomial connection coefficients*, which are given as

$$\Gamma(j_1, \vec{k}_1, \vec{m}_1, \dots, j_g, \vec{k}_g, \vec{m}_g, \vec{l}) = \int_{\mathbb{R}^n} \vec{x}^{\vec{l}} \cdot \prod_{i=1}^g \mathbf{D}_{\vec{m}_i} \varphi_{j_i, \vec{k}_i}(\vec{x}) d\vec{x}.$$

The calculation of such generalized connection coefficients is analogous to the way presented for simple connection coefficients. They mainly rely on the

refinable properties of the scaling functions and wavelets as well as on their vanishing moments.

EXAMPLE* 6 We shall calculate the connection coefficients for first order differentiation in an explicit example now. Once again, we consider the quincunx dilation matrix

$$\mathbf{Q} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

One directly sees that $\mathbf{Q}^2 = 2 \cdot \mathbf{Id}$. For $\vec{m} = [1, 0]^T$, the relations (6.10) become

$$\begin{aligned} \Gamma_{\varphi, [1, 0]^T}^{\varphi}(0, \vec{0}, 0, \vec{k}, 0) &= \sum_{\vec{l}} \sum_{\vec{q}} h_{\vec{l}} \cdot h_{\vec{q}} \cdot \Gamma_{\varphi, [1, 0]^T}^{\varphi}(0, \vec{0}, 0, \mathbf{Q} \cdot \vec{k} - \vec{l} + \vec{q}, 1), \\ \Gamma_{\varphi, [1, 0]^T}^{\varphi}(0, \vec{0}, 0, \vec{k}, 1) &= 2 \cdot \sum_{\vec{l}} \sum_{\vec{q}} h_{\vec{l}} \cdot h_{\vec{q}} \cdot \Gamma_{\varphi, [1, 0]^T}^{\varphi}(0, \vec{0}, 0, \mathbf{Q} \cdot \vec{k} - \vec{l} + \vec{q}, 0), \end{aligned}$$

while equations (6.11) turn into

$$\begin{aligned} \sum_{\vec{k}} \vec{k}^{[1, 0]^T} \Gamma_{\varphi, [1, 0]^T}^{\varphi}(0, \vec{0}, 0, \vec{k}, 0) &= -1, \\ \sum_{\vec{k}} \vec{k}^{[0, 1]^T} \Gamma_{\varphi, [1, 0]^T}^{\varphi}(0, \vec{0}, 0, \vec{k}, 0) &= 0. \end{aligned}$$

If we exemplarily apply the POWELL-ZWART scaling filter mask

$$h_{[0, 0]} = h_{[1, 0]} = h_{[0, 1]} = h_{[1, 1]} = \frac{\sqrt{2}}{4},$$

to this situation, the fundamental connection coefficients with respect to the exponent $[1, 0]^T$ (this means the operator \mathbf{D}_{x_1}) are calculated as

$$\Gamma_{\varphi, [1, 0]^T}^{\varphi} = \begin{bmatrix} 0 & \frac{1}{240} & 0 & -\frac{1}{240} & 0 \\ \frac{1}{160} & \frac{23}{240} & 0 & -\frac{23}{240} & -\frac{1}{160} \\ \frac{7}{240} & \frac{13}{60} & \mathbf{0} & -\frac{13}{60} & -\frac{7}{240} \\ \frac{1}{160} & \frac{23}{240} & 0 & -\frac{23}{240} & -\frac{1}{160} \\ 0 & \frac{1}{240} & 0 & -\frac{1}{240} & 0 \end{bmatrix}.$$

Hereby, the boldface printed zero indicates the coordinate origin.

6.1.3. Relations to Finite Difference Schemes

In the previous paragraph, we have shown that the fundamental connection coefficients (and thus *all* connection coefficients) are calculable under a certain mild criterion regarding the scaling matrix \mathbf{Q} . Assuming now, that we have a set of connection coefficients at hand, we may return to the question of the evaluation of functional derivatives. If we insert a given set of connection coefficients into the representation (6.3), we obtain

$$\begin{aligned}
 \mathbf{D}_{\vec{m}}(f)(\vec{x}) \approx & \sum_{\vec{k}} \sum_{\vec{l}} c_{\vec{k}}^i \cdot \Gamma_{\varphi, \vec{m}}^{\varphi}(j, \vec{k}, j, \vec{l}, 0) \cdot \varphi_{j, \vec{k}}(\vec{x}) + \\
 & \sum_{p=1}^{|q|-1} \sum_{i=1}^j \sum_{\vec{k}} \sum_{\vec{l}} d_{\vec{k}}^{i,p} \cdot \Gamma_{\psi^p, \vec{m}}^{\varphi}(i, \vec{k}, j, \vec{l}, 0) \cdot \varphi_{j, \vec{k}}(\vec{x}) + \\
 & \sum_{p=1}^{|q|-1} \sum_{i=1}^j \sum_{\vec{k}} \sum_{\vec{l}} c_{\vec{k}}^j \cdot \Gamma_{\varphi, \vec{m}}^{\psi^p}(j, \vec{k}, i, \vec{l}, 0) \cdot \psi_{i, \vec{k}}^p(\vec{x}) + \\
 & \sum_{p=1}^{|q|-1} \sum_{r=1}^{|q|-1} \sum_{i=1}^j \sum_{\vec{k}} \sum_{\vec{l}} d_{\vec{k}}^{i,p} \cdot \Gamma_{\psi^p, \vec{m}}^{\psi^r}(i, \vec{k}, i, \vec{l}, 0) \cdot \psi_{i, \vec{k}}^p(\vec{x}).
 \end{aligned} \tag{6.12}$$

This expression may look rather complicated at first sight, but a closer look shows that these expressions are just discrete convolutions between the data $(c_{\vec{k}}^i, d_{\vec{k}}^{i,p})$ and the connection coefficients if the scales coincide, which means that they can be computed in linear time. Between different scales, the evaluation becomes a *dilated convolution* that is only feasible in log-linear time, which means that data described as a multiscale decomposition requires more computational amount (but on the other hand less storage). However, later in this section we will present a method to evaluate the full operators even for multiscale decomposed data in linear time. Anyway, this representation makes it evident that the connection coefficients are in fact the discrete transformation kernel for the differential operator $\mathbf{D}_{\vec{m}}$ in the wavelet basis. And due to the rapid decay of wavelets, only very few connection coefficients are significantly away from zero, which means that $\mathbf{D}_{\vec{m}}$ is very narrow-banded in the wavelet domain (recall the remarks at the beginning of this chapter) giving a very efficient way of calculation of partial derivatives.

Connection coefficients share the property of giving approximative representations of functional derivatives by discrete convolutions with classical finite difference schemes like e.g. the SOBEL operator or more sophisticated schemes originating from STIRLINGS approximation formula. By using the commuta-

tion rule

$$\mathbf{D}_{\vec{m}} \circ \mathbf{P}_j(f) = \mathbf{P}_j \circ \mathbf{D}_{\vec{m}}(f), \quad (6.13)$$

we can directly give the following estimate for the approximation quality of the wavelet based differentiation.

COROLLARY* 6.8 *Given a \mathbf{Q} -adic MRA built from a scaling function $\varphi(\vec{x})$ and $|q| - 1$ wavelets $\psi^p(\vec{x})$ such that the connection coefficients are known and*

$$\mu_{\vec{l}}^{\psi^p} = 0, \quad p = 1, \dots, |q| - 1$$

for all \vec{l} with $|\vec{l}| \leq \nu \in \mathbb{N}$. Then, for every compactly supported function f having at least ν continuous derivatives and all \vec{m} with $|\vec{m}| \leq \nu$, we obtain

$$\|\mathbf{D}_{\vec{m}} \circ \mathbf{P}_j(f) - \mathbf{D}_{\vec{m}}(f)\|_{L_2} \leq C \cdot |\det \mathbf{Q}|^{j \cdot (\nu - |\vec{m}|)}. \quad (6.14)$$

Proof. Follows directly from (6.13) and Proposition 5.2. \square .

This approximation order coincides with those achieved by STIRLING operators in the onedimensional case for first order derivatives and gives thus an optimal exponential decay of the approximation error. However, since the design of multidimensional filters with optimal zero orders (giving the required vanishing moments) leaves some degrees of freedom for further design features, one can additionally optimize the connection coefficients in such a way that they also give an optimized approximation error with respect to other criteria. For example, in image analysis it is of special interest to have filters that minimize the angular errors of gradient vectors; the usefulness of such filters as well as their derivation was first demonstrated in [83] and further improved in [200]. Maintaining the remaining degrees of freedom in the wavelet filter design procedure, we can do the same yielding competitive results.

EXAMPLE* 7 *Suppose that we want to optimize the connection coefficients for a symmetric 4×4 filter $[\mathbf{h}_{\vec{k}}]_{\vec{k} \in \mathbb{Z}^2}$ that belongs to a scaling function with the regular dilation matrix $\mathbf{Q} = 2 \cdot \mathbf{Id}$ with respect to the isotropy criteria developed in [200]. To guarantee some polynomial reproducibility, the filter may be equipped with one vanishing moment which (after normalization) yields*

$$\mathbf{h} = \begin{bmatrix} \frac{1}{2} - 2\eta - \gamma & \eta & \eta & \frac{1}{2} - 2\eta - \gamma \\ \eta & \gamma & \gamma & \eta \\ \eta & \gamma & \gamma & \eta \\ \frac{1}{2} - 2\eta - \gamma & \eta & \eta & \frac{1}{2} - 2\eta - \gamma \end{bmatrix}$$

with free parameters η and γ that will be used for our optimization purposes. Calculating the connection coefficients for first order differentiation for the filter $[h_k]_{\vec{k} \in \mathbb{Z}^2}$ leads to a coefficient set given by rational functions in η and γ . The angular error that is to minimize is given by

$$\epsilon = \left| \arctan \frac{\hat{\Gamma}_{\varphi,[0,1]^T}^{\varphi}(\vec{k})}{\hat{\Gamma}_{\varphi,[1,0]^T}^{\varphi}(\vec{k})} - \arctan \frac{k_2}{k_1} \right| \quad (\text{see [200]}),$$

where $\hat{\Gamma}_{\varphi,[\cdot,\cdot]^T}^{\varphi}(\vec{k})$ denotes the FOURIER transform of the connection coefficient mask. A by-hand minimization of this expression leads to the near-optimal (dyadic rational) values $\eta = \frac{27}{256}$ and $\gamma = \frac{65}{256}$, which give (approximately) the connection coefficients

$$\Gamma_{\varphi,[1,0]^T}^{\varphi} \approx \begin{bmatrix} 0.00050205 & 0.00459532 & 0 & -0.00459532 & -0.00050205 \\ 0.01157092 & 0.08981441 & 0 & -0.08981441 & -0.01157092 \\ 0.02831520 & 0.20625825 & 0 & -0.20625825 & -0.02831520 \\ 0.01157092 & 0.08981441 & 0 & -0.08981441 & -0.01157092 \\ 0.00050205 & 0.00459532 & 0 & -0.00459532 & -0.00050205 \end{bmatrix}$$

and

$$\Gamma_{\varphi,[0,1]^T}^{\varphi} = \Gamma_{\varphi,[1,0]^T}^{\varphi T}.$$

These connection coefficients give indeed very well-performing derivative filters in 2D. In Figure 6.1, the magnitude of the absolute angular error is compared to the results achieved by two other commonly used derivative filters.

We conclude that the optimized connection coefficients do not only perform even slightly better than the optimized filters from [200], but they also can be applied extremely efficiently directly in the multiscale wavelet domain (i.e. without back-transformation), while other derivative operators can only be applied on the discrete data in single-scale representation.

The reader may ask why derivative filters derived from connection coefficients can perform better than *optimized* derivative filters — aren't they really optimized eventually? The simple answer to this questions is: they *are* optimized, but they are built in a separable manner. It is again the nonseparability of the applied wavelet filters that makes further improvements possible.

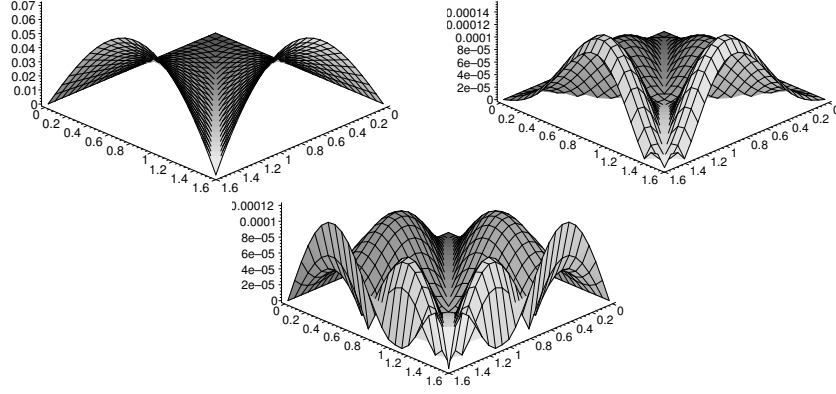


Figure 6.1. Absolute values of the angular errors. The reader should note the different axis scalings. *Upper Left.* SOBEL filter. *Upper Right.* Optimized 5×5 filter from [200]. *Lower Row.* Optimized 5×5 connection coefficient mask.

6.2. Affine Transforms in Wavelet Domain

In this section we will investigate operators for affine transforms in the wavelet domain. Such transformations are of special interest in image sequence analysis, since motion in real-world scenes is physically modelled by translation and scaling in the image plane. In our considerations, we will mainly care about translation operators, that is

$$\mathbf{T}_{\vec{\tau}} : f(\vec{x}) \longrightarrow f(\vec{x} - \vec{\tau}(\vec{x})) \quad (6.15)$$

with a *translation field* $\vec{\tau} : \mathbb{R}^n \rightarrow \mathbb{R}^n$. To stay consistent with the modelling done so far, we will only use discretized translations, which means that the translation field is given by a discrete array $[\vec{\tau}_{\vec{k}}]_{\vec{k} \in \Omega} \in \mathbb{R}^{|\Omega|}$, where Ω denotes the data range. Recalling the idealized assumption (5.12), the discrete version of (6.15) in the wavelet domain reads

$$\mathbf{T}_{\vec{\tau}} : f(\vec{x}) = \sum_{\vec{k} \in \mathbb{Z}^n} c_{\vec{k}}^0 \cdot \varphi_{0,\vec{k}}(\vec{x}) \longrightarrow \sum_{\vec{k} \in \mathbb{Z}^n} c_{\vec{k}}^0 \cdot \varphi_{0,\vec{k}}(\vec{x} - \vec{\tau}_{\vec{k}}). \quad (6.16)$$

Since an efficient processing requires to have a corresponding operator at a multiscale-decomposed data set at hand, one obtains under usage of (5.14) and (5.15)

$$\begin{aligned} \mathbf{T}_{\vec{\tau}} : f(\vec{x}) \longrightarrow & \sum_{\vec{k} \in \mathbb{Z}^n} c_{\vec{k}}^i \cdot \varphi_{i,\vec{k}}(\vec{x} - \mathbf{Q}^i \cdot \vec{\tau}_{\vec{k}}) + \\ & \sum_{p=1}^{|q|-1} \sum_{j=1}^i \sum_{\vec{k} \in \mathbb{Z}^n} d_{\vec{k}}^{j,p} \cdot \psi_{j,\vec{k}}^p(\vec{x} - \mathbf{Q}^j \cdot \vec{\tau}_{\vec{k}}). \end{aligned} \quad (6.17)$$

It is evident that the fundamental need is now to find a simple and efficient way to get a representation

$$\mathbf{T}_{\vec{\tau}}(f)(\vec{x}) = \sum_{\vec{k} \in \mathbb{Z}^n} \alpha_{\vec{k}}^i \cdot \varphi_{\vec{i}, \vec{k}}(\vec{x}) + \sum_{p=1}^{|q|-1} \sum_{j=1}^i \sum_{\vec{k} \in \mathbb{Z}^n} \beta_{\vec{k}}^{j,p} \cdot \psi_{j, \vec{k}}^p(\vec{x}) \quad (6.18)$$

from (6.17). Again, the problem essentially reduces to evaluate the expansion coefficients

$$\begin{aligned} \alpha_{\vec{k}}^i &= \sum_{\vec{l} \in \mathbb{Z}^n} c_{\vec{l}}^i \cdot \int_{\mathbb{R}^n} \varphi_{i, \vec{k}}(\vec{x} - \mathbf{Q}^i \cdot \vec{\tau}_{\vec{k}}) \cdot \varphi_{i, \vec{l}}(\vec{x}) d\vec{x} + \\ &\quad \sum_{p=1}^{|q|-1} \sum_{j=1}^i \sum_{\vec{l} \in \mathbb{Z}^n} d_{\vec{l}}^{j,p} \cdot \int_{\mathbb{R}^n} \psi_{j, \vec{k}}^p(\vec{x} - \mathbf{Q}^j \cdot \vec{\tau}_{\vec{k}}) \cdot \varphi_{i, \vec{l}}(\vec{x}) d\vec{x} \end{aligned} \quad (6.19)$$

and $\beta_{\vec{k}}^{j,p}$ which is given by a similar expression. Obviously, the inner products in (6.19) can be evaluated exactly for all \mathbf{Q} -adic rational numbers by (probably many) repeated applications of the scaling equation. However, this proceeding gets very lengthy in common and is not fitted to deal with non- \mathbf{Q} -adic rational shifts. In this situation, it is preferable to approximate the inner products by suited quadrature formulae.¹ Hereby, the connection coefficients introduced in the previous section will be an important helping tool. We have to consider the integrals

$$t(\vec{\tau}) = \int_{\mathbb{R}^n} \varphi(\vec{x}) \cdot \varphi(\vec{x} - \vec{\tau}) d\vec{x},$$

where $\vec{\tau} \in \mathbb{R}^n$ is an arbitrary vector (possibly of not too large absolute value). The more complicated integral expressions in (6.19) can be reduced to integrals of this simple type under usage of the refinable properties of φ (as in the previous section). Returning to the question of the efficient approximative calculation of $t(\vec{\tau})$, we choose $\vec{\zeta} \in \mathbb{Z}^n$ to be the integer vector which is closest to $\vec{\tau}$ in the EUCLIDEAN norm, i.e. $\|\vec{\tau} - \vec{\zeta}\|_2 = \min!$. Then, a TAYLOR expansion leads to

$$\begin{aligned} t(\vec{\tau}) &= \int_{\mathbb{R}^n} \varphi(\vec{x}) \cdot \varphi(\vec{x} - \vec{\tau}) d\vec{x} \\ &= \int_{\mathbb{R}^n} \varphi(\vec{x}) \cdot \left(\sum_{\vec{m} \in \mathbb{N}_0^n} \frac{1}{\vec{m}!} \cdot (\vec{\zeta} - \vec{\tau})^{\vec{m}} \cdot \frac{\partial^{|\vec{m}|}}{\partial \vec{x}^{\vec{m}}} \varphi(\vec{x} - \vec{\zeta}) \right) d\vec{x} \end{aligned}$$

¹There may be exceptions to this situation. For example, under certain conditions, refinable functions may coincide with splines and for such piecewise polynomial functions, expressions like $\int_{\mathbb{R}^n} \varphi(\vec{x}) \cdot \varphi(\vec{x} - \vec{\tau}) d\vec{x}$ can be evaluated exactly, but this is not true in general.

and cutting off the TAYLOR series yields an approximation²

$$\begin{aligned} t(\vec{\tau}) &\approx \int_{\mathbb{R}^n} \varphi(\vec{x}) \cdot \left(\sum_{|\vec{m}| \leq \nu} \frac{1}{\vec{m}!} \cdot (\vec{\zeta} - \vec{\tau})^{\vec{m}} \cdot \frac{\partial^{|\vec{m}|}}{\partial \vec{x}^{\vec{m}}} \varphi(\vec{x} - \vec{\zeta}) \right) d\vec{x} \\ &= \sum_{|\vec{m}| \leq \nu} \frac{1}{\vec{m}!} \cdot (\vec{\zeta} - \vec{\tau})^{\vec{m}} \cdot \Gamma_{\varphi, \vec{m}}^{\varphi}(0, \vec{\zeta}, 0, \vec{0}, 0), \end{aligned} \quad (6.20)$$

that is expressed by a weighted number of fundamental connection coefficients. In a completely analogous manner, all the expansion coefficients from (6.19) can be calculated directly as weighted sums of higher order connection coefficients giving us the desired representation of the translated data as a wavelet series.

This treatment already implies the announced fast algorithm for image warping, which is nothing else but calculating a translated image representation for a given translation field $\vec{\tau}$. The algorithm is very fast (linear time for single-scale and log-linear time for multiscale representations) since for every data point (pixel) only a fixed number of multiplications of shifts with connection coefficients have to be performed, followed by a discrete filtering with the expansion coefficients c_k^i and $d_k^{j,p}$ of the original data by (6.19).

We conclude this section with some remarks about other operators, which have some significance in image processing. The first such class are scaling operators, that can be approximated in an analogous way as translation operators, with the slight difference that in this situation, integrals of the type

$$\sigma(\Lambda_{\vec{k}}) = \int_{\mathbb{R}^n} \varphi(\vec{x}) \cdot \varphi(\Lambda_{\vec{k}} \cdot \vec{x} - \vec{k}) d\vec{x}$$

with scaling matrices $\Lambda_{\vec{k}} \in \text{Mat}(n \times n, \mathbb{R})$ have to be approximated. This again requires the usage of connection coefficients as well as polynomial connection coefficients. However, we will not need scaling operators for the applications in this thesis and consequently, we will not go further into the details here.

Another transform of importance in image analysis is the HILBERT transform and its multidimensional generalizations, the RIESZ transforms \mathbf{R}_j . A representation of the HILBERT transform in wavelet coordinates was developed in [22]. This construction is again similar to that for connection coefficients and

²In practice, we will cut off the TAYLOR series after the first or second order derivatives, which gives reliable results for translations of not too big absolute values.

is easily extendible to its n -dimensional generalizations of RIESZ transforms. Their description in the wavelet domain is given via the integrals

$$r_j(\vec{k}) = \lim_{\epsilon \rightarrow 0} \int_{\mathbb{R}^n} \int_{|\vec{y}| \geq \epsilon} \varphi(\vec{x} - \vec{k}) \cdot \varphi(\vec{x} - \vec{y}) \cdot \frac{\Gamma(\frac{n+1}{2})}{\pi^{(n+1)/2}} \cdot \frac{y_j}{|\vec{y}|^{n+1}} d\vec{y} d\vec{x}.$$

Following [22], this discrete transformation kernel can be evaluated up to arbitrary precision under exploitation of the asymptotics for the $r_j(\vec{k})$. These asymptotics are, under the assumption that the applied wavelets have a certain number of vanishing moments, given by

$$r_j(\vec{k}) \longrightarrow \frac{\Gamma(\frac{n+1}{2})}{\pi^{(n+1)/2}} \cdot \frac{k_j}{|\vec{k}|^{n+1}}$$

for large absolute values of \vec{k} . This can be easily verified by a series expansion.

6.3. Two Ways to Represent Operators in Wavelet Coordinates

Representations of operators in the wavelet domain are feasible in more than one way. A description as e.g. in (6.12) was straightforward obtainable via the operators $\mathbf{P}_j(f)$ and $\mathbf{U}_j(f)$. This is called the *standard form* of representation, but we had to state that their execution is only feasible in log-linear time as long as the data is given in the sparser multiscale representation; this will be briefly concluded in the first paragraph of this section. In the second paragraph, we will introduce a generalization first given in [20] leading to the *non-standard form* of representation yielding linear time algorithms. Following our intentions for this thesis, we will present this in a generalized form suited for multidimensional data and matrix dilations with not necessarily dyadic subsampling.

6.3.1. The Standard Form

Suppose that we are given any operator \mathbf{T} for which we want to give a wavelet-based description. Assuming that there exists a finest scale, say $j = 0$, the operator \mathbf{T} satisfies

$$\mathbf{T} = \mathbf{P}_0 \mathbf{T} \mathbf{P}_0 = \mathbf{P}_0 \mathbf{T} \mathbf{P}_j + \sum_{i=1}^j \sum_{m=1}^{|\mathbf{q}|-1} \mathbf{P}_0 \mathbf{T} \mathbf{U}_i^m.$$

The second equality follows from the defining relation (5.5). Repeated application of this procedure leads to a representation

$$\mathbf{T} = \mathbf{P}_j \mathbf{T} \mathbf{P}_j + \sum_{i=1}^j \sum_{k=1}^j \sum_{m=1}^{|\mathbf{q}|-1} \sum_{p=1}^{|\mathbf{q}|-1} \mathbf{U}_k^p \mathbf{T} \mathbf{U}_i^m. \quad (6.21)$$

Reviewing the example of differentiation, the projections $\mathbf{P}_j \mathbf{T} \mathbf{P}_j$ and $\mathbf{U}_k^p \mathbf{T} \mathbf{U}_i^m$ are described by the different connection coefficients. Writing \mathbf{T} via (6.21) in matrix notation, one obtains a sparse matrix $\tilde{\mathbf{T}}$ whose entries are the related connection coefficients, that looks like in Figure 6.2.

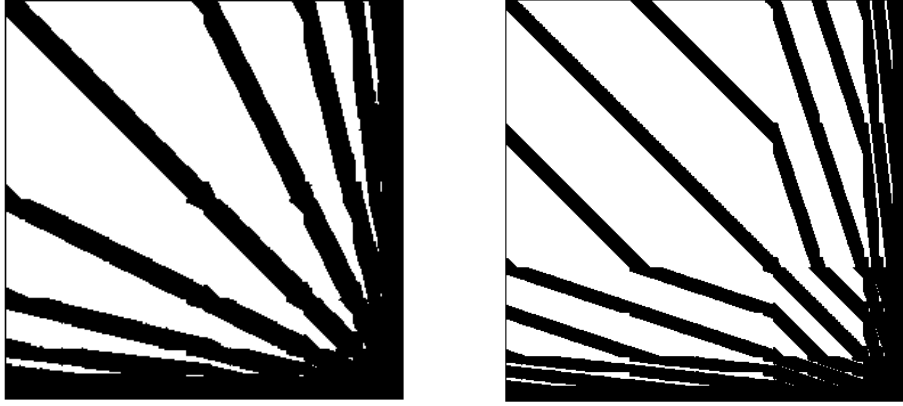


Figure 6.2. Typical sparsity patterns for matrices associated to an operator in the standard form. *Left.* Dyadic wavelets — 1D example with data length $N = 256$ giving an operator matrix $\tilde{\mathbf{T}}$ of size 256×256 . *Right.* Triadic wavelets — 1D example of length $N = 729$.

This means that at least for orthogonal wavelets, several operators \mathbf{T} acting on a discrete data set of size $N = |q|^{\nu \cdot n}$ can be efficiently represented by a sparse matrix $\tilde{\mathbf{T}} \in \text{Sparse}(|q|^{\nu \cdot n} \times |q|^{\nu \cdot n}, \mathbb{R})$.³ It is easily proved that only $O(N \cdot \nu \cdot n)$ entries of the matrix are significantly larger than zero and consequently, the numerical cost for the evaluation of the matrix-data product yielding the operation can be essentially done in log-linear time.

6.3.2. The Non-Standard Form

Representation (6.21) is the straightforward way to describe operators in the wavelet domain. This leads to sparse and hierarchical structures yielding fast transformation algorithms. However, the representation can be made even sparser by *decoupling the scales*, that is, finding a decomposition that does not mix information from different scales as it is the case in (6.21). The key to this

³ Again, the reader should note that the same is true for a tight wavelet frame. But since the overdetermination of such a system, the matrix $\tilde{\mathbf{T}}$ is of larger size.

is the operator decomposition (5.3) and the usage of a telescopic sum giving

$$\begin{aligned}
 \mathbf{T} &= \mathbf{P}_0 \mathbf{T} \mathbf{P}_0 = \mathbf{P}_1 \mathbf{T} \mathbf{P}_0 + \sum_{k=1}^{|q|-1} \mathbf{U}_1^k \mathbf{T} \mathbf{P}_0 \\
 &= \mathbf{P}_j \mathbf{T} \mathbf{P}_j + \sum_{i=1}^j \sum_{k=1}^{|q|-1} \left(\mathbf{U}_i^k \mathbf{T} \mathbf{P}_i + \mathbf{P}_i \mathbf{T} \mathbf{U}_i^k + \sum_{l=1}^{|q|-1} \mathbf{U}_i^k \mathbf{T} \mathbf{U}_i^l \right). \tag{6.22}
 \end{aligned}$$

Obviously, the goal to completely separate the scales is achieved by this decomposition. But one should also note, that the representation requires low pass information on all scales (not only on the coarsest as in a *usual* wavelet representation) due to the presence of the projectors \mathbf{P}_i for all $i = 1, \dots, j$, which means, that this operator representation acts on an overdetermined data representation. This fact causes some consequences, namely one is urged to preserve all expansion coefficients c_k^i during a wavelet decomposition, the describing operator matrix $\tilde{\mathbf{T}}$ becomes larger than in the standard form and a suited back-transformation into the *original wavelet domain* is necessary after the operator is executed. To give a short example, the derivative operator $\mathbf{D}_{\vec{m}}$, whose standard form for a wavelet-approximated function $f(\vec{x})$ is (6.12), is given by

$$\begin{aligned}
 \mathbf{D}_{\vec{m}}(f)(\vec{x}) &\approx \sum_{\vec{k}} \sum_{\vec{l}} c_{\vec{k}}^i \cdot \Gamma_{\varphi, \vec{m}}^{\varphi}(j, \vec{k}, j, \vec{l}, 0) \cdot \varphi_{j, \vec{k}}(\vec{x}) + \\
 &\sum_{p=1}^{|q|-1} \sum_{i=1}^j \sum_{\vec{k}} \sum_{\vec{l}} d_{\vec{k}}^{i,p} \cdot \Gamma_{\psi^p, \vec{m}}^{\varphi}(i, \vec{k}, i, \vec{l}, 0) \cdot \varphi_{i, \vec{k}}(\vec{x}) + \\
 &\sum_{p=1}^{|q|-1} \sum_{i=1}^j \sum_{\vec{k}} \sum_{\vec{l}} c_{\vec{k}}^i \cdot \Gamma_{\varphi, \vec{m}}^{\psi^p}(i, \vec{k}, i, \vec{l}, 0) \cdot \psi_{i, \vec{k}}^p(\vec{x}) + \\
 &\sum_{p=1}^{|q|-1} \sum_{r=1}^{|q|-1} \sum_{i=1}^j \sum_{\vec{k}} \sum_{\vec{l}} d_{\vec{k}}^{i,p} \cdot \Gamma_{\psi^p, \vec{m}}^{\psi^r}(i, \vec{k}, i, \vec{l}, 0) \cdot \psi_{i, \vec{k}}^p(\vec{x}) \tag{6.23}
 \end{aligned}$$

in the non-standard form. This is now obviously a genuine convolution operator and thus feasible in linear time depending on the data size. It is an easy task to verify that the describing matrix $\tilde{\mathbf{T}}$ in the non-standard form for a data set of size $N = |q|^{\nu \cdot n}$ is in

$$\tilde{\mathbf{T}} \in \text{Sparse} \left(|q| \cdot \frac{|q|^{\nu \cdot n} - 1}{|q| - 1} \times |q| \cdot \frac{|q|^{\nu \cdot n} - 1}{|q| - 1}, \mathbb{R} \right),$$

which is between $1 + \epsilon$ and 4 times larger than in the standard form. But the matrix has a more sparse and stricter hierarchical structure, allowing it to be

executed in $O(N)$ significant operations [20]. A typical example how the non-standard form matrices look like is given in Figure 6.3.

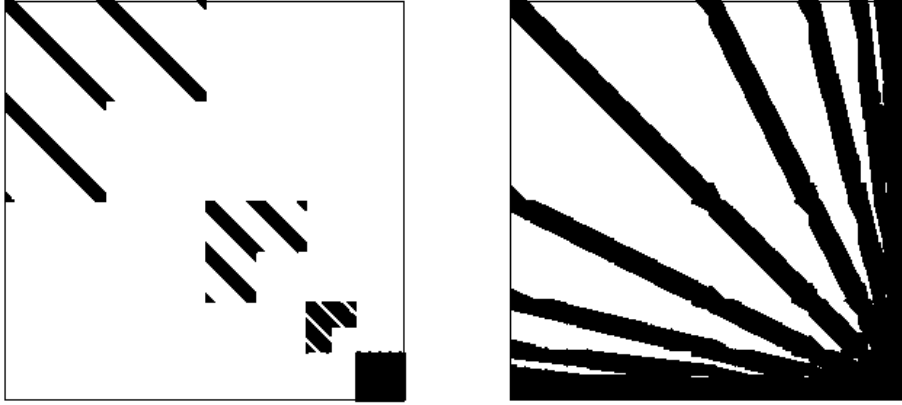


Figure 6.3. Typical sparsity pattern for matrices associated to an operator in the non-standard form. *Left.* Dyadic wavelets — 1D example with data length $N = 256$ giving an operator matrix \tilde{T} of size 510×510 . *Right.* In comparison the corresponding $N \times N = 256 \times 256$ standard form representation as shown in the left illustration of Figure 6.2. Obviously, the non-standard form is vastly sparser.

We conclude that both operator representations have their advantages as well as disadvantages and it is not to say that one of the both is preferable in general. Even the faster execution of the non-standard form might be valueless when the data size is not too large and an extra back-transformation into the wavelet domain is necessary for further purposes. On the other hand, the more compact representation of the standard form might become problematical even for smaller data sizes if one is interested in efficient inversions. In practice, the decision between the both approaches should depend on a balancing of the different requirements.

6.4. Related Work

Not only differential and affine operators as presented in the first parts of this chapter possess sparse and diagonally dominant representations in the wavelet domain. This is in fact true for a much wider class of operators, which are contained in the classical algebra of operators of CALDERÓN-ZYGMUND type [36] [37] as well as for so-called *pseudo-differential* operators. The investigation of the properties of such operators and their description in the wavelet domain requires several deeper results from harmonic analysis and lies beyond the purposes of this thesis. The interested reader will find lots of material in

the books of MEYER and COIFMAN [157] [159] as well as in the lecture notes [118]. However, to give the reader an idea about the magnitude of the class of operators that have sparse and simple representations in wavelet coordinates, we quote the following result.

THEOREM 6.9 *Given a scaling function $\varphi(\vec{x})$ and an associated set of $|q| - 1$ wavelets $\psi^p(\vec{x})$ having at least ν vanishing moments and an operator \mathbf{T} with transformation kernel $k(\vec{x}, \vec{y})$ such that*

$$\begin{aligned} |k(\vec{x}, \vec{y})| &\leq \frac{C_0}{|\vec{x} - \vec{y}|} \quad \text{and} \\ |\mathbf{D}_{\vec{x}\vec{m}} k(\vec{x}, \vec{y})| + |\mathbf{D}_{\vec{y}\vec{m}} k(\vec{x}, \vec{y})| &\leq \frac{C_{\vec{m}}}{|\vec{x} - \vec{y}|^{1+|\vec{m}|}}. \end{aligned}$$

Under the additional condition, that either \mathbf{T} is uniformly bounded or satisfies (at least) the weak cancellation property

$$\left| \int_{\Omega \times \Omega} k(\vec{x}, \vec{y}) \, d\vec{x} \, d\vec{y} \right| \leq C \cdot \text{vol}(\Omega)$$

for all \mathbf{Q} -adic subsets $\Omega \subset \mathbb{R}^n$, the following estimate holds true

$$\sum_{r=0}^{|q|-1} \sum_{p=0}^{|q|-1} |\mu_{\vec{k}, \vec{l}}^{j,p,r}| \leq \frac{C_\nu}{1 + |\vec{k} - \vec{l}|^{\nu+1}}. \quad (6.24)$$

Hereby, the $\mu_{\vec{k}, \vec{l}}^{j,p,r}$ are the connection coefficients describing the operator in the wavelet domain, in particular

$$\mu_{\vec{k}, \vec{l}}^{j,p,r} = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} k(\vec{x}, \vec{y}) \cdot \psi_{j,\vec{k}}^p(\vec{x}) \cdot \psi_{j,\vec{l}}^r(\vec{y}) \, d\vec{x} \, d\vec{y}.$$

Of course, the estimate (6.24) allows one to drop all those coefficients $\mu_{\vec{k}, \vec{l}}^{j,p,r}$ below a significance threshold ϵ and thus yields a sparse non-standard representation of \mathbf{T} in the wavelet domain. In this sense, Theorem 6.9 gives indeed a characterization of all those operators having an efficient description via wavelet transformations.

The *proof* of Theorem 6.9 is found in [20] for the 1D case. Since its n -dimensional generalization is straightforward, it is omitted here. We finally quote that similar results were also obtained in [219]. \square .

Chapter Summary

The actual chapter dealt with the representation of operators in wavelet bases. Hereby, connection coefficients turned out to be the key concept, since it was

demonstrated that they yield a discrete finger-banded matrix representation of the considered operator. We have shown that connection coefficients are always calculable under a mild condition regarding the involved dilation matrix. Out of this result, we have exemplarily presented a way to calculate connection coefficients associated to differential operators and have shown their close relation to other finite difference schemes. A result about the approximation quality for this type of wavelet-represented operators was also given. Two examples showing the potential of such differentiation schemes followed afterwards. In the second section, a framework for the representation of affine operators, which play a prominent role in image analysis, was sketched (the details were left out due to their similarity with the differential connection coefficients). The chapter ended by recalling some older results about the different possible wavelet-based matrix representations of onedimensional operators and the (straightforward) generalization to the higherdimensional situation was also included there.

Chapter 7

REGULARIZATION AND SOLUTION OF ILL-POSED PROBLEMS

Es lassen sich positive Zahlen ρ und ϵ finden, so daß bei einem etwaigen Zusammenstoß von Erde und Sonne der Mond mindestens den Abstand ρ von der Erde hat und dann noch mindestens die Zeit ϵ braucht, bis auch er vielleicht mit ihr zusammenstößt. Dieses Ergebnis wird uns dabei helfen, zuversichtlich in die Zukunft zu schauen

—CARL LUDWIG SIEGEL [207]

This chapter is dedicated to develop efficient wavelet-based methods to solve various mathematically ill-posed problems. In the first section, we will give a mathematical formulation of a model problem, that will be fundamental for all further investigations. Furthermore, the main method we will use in applications, namely the *GALERKIN projection method*, will be introduced. The second part of this chapter is concerned with some considerations about the numerical realization of solving ill-posed problems. Hereby, the interest is first put on preconditioning of the equation systems and on time discretization schemes; afterwards, we will develop a strategy to reduce the dimensional complexity and increase the stability and convergence of the numerical procedure to solve the original problem. Most of the results in this chapter are well-known to numerical analysts, but we recall them here for the sake of completeness, for assuring that they also work in our generalized setting of multidimensional, matrix dilation multiresolution analyses and for combining them in a new fashion.

7.1. Problem Statement

In many scientific applications, the problem occurs to recover the initial or an intermediate state of some *system* from a set of measured incidents. Mathematically, such a system can be usually formulated by means of (not necessarily linear) operators of integro-differential type, while the measured incidents as

well as the systems' intermediate states are modelled as functions of time and/or space. So, the problem may be written as the search for a function $f(\vec{x})$ that satisfies

$$\mathbf{T}(f)(\vec{x}) = g(\vec{x}) \quad (7.1)$$

where the operator \mathbf{T} and the measurement(s) $g(\vec{x})$ are assumed to be known. Sometimes, different types of additional boundary conditions (e.g. of NEUMANN or DIRICHLET type) are also to be satisfied. The difficulty that often occurs is, that there exists no continuous inverse of \mathbf{T} and that the measurement(s) are usually polluted by some kind of noise. Therefore, numerical approximations are needed to establish some regularity and soundness in this situation. Problems of this kind are usually called *inverse problems* or *ill-posed problems*.¹

7.1.1. The Galerkin Method in the Wavelet Domain

The fundamental idea of the GALERKIN method is to project problems like (7.1) onto local subspaces instead of considering it globally. Hereby, the nested MRA decomposition of the signal space via the wavelet transform is a very attractive choice. It is well-known from linear algebra that the orthogonal projection of a vector \vec{w} onto a subspace V is given by taking the inner product of \vec{w} with the basis elements spanning V . In terms of a wavelet decomposition, this is maintained by the operators \mathbf{P}_j and \mathbf{U}_j^m defined in Chapter 5. The wavelet GALERKIN formulation of (7.1) then reads

$$\mathbf{P}_j \mathbf{T}(f)(\vec{x}) + \sum_{k \leq j} \sum_{m=1}^{q-1} \mathbf{U}_k^m \mathbf{T}(f)(\vec{x}) = \mathbf{P}_j(g)(\vec{x}) + \sum_{k \leq j} \sum_{m=1}^{q-1} \mathbf{U}_k^m(g)(\vec{x}). \quad (7.2)$$

Hereby, the wavelet representation $\mathbf{P}_j \mathbf{T}$ of the operator \mathbf{T} should follow the guidelines developed in the previous chapter. This gives a description of the original problem in terms of local subspaces, but the statement itself remains global. To *test* the validity of (7.2) locally, one has to take the inner products of (7.2) with the local scaling functions $\varphi_{j,\vec{k}}(\vec{x})$ and wavelets $\psi_{j,\vec{k}}^p(\vec{x})$ yielding

¹In the sense of HADAMARD, a problem is called *well-posed* if a unique solution exists and this solution depends continuously on the initial data. Otherwise, it is called *ill-posed*.

an equation system

$$\left\langle \mathbf{P}_j \mathbf{T}(f)(\vec{x}) + \sum_{k \leq j} \sum_{m=1}^{q-1} \mathbf{U}_k^m \mathbf{T}(f)(\vec{x}), \varphi_{j,\vec{k}}(\vec{x}) \right\rangle = \left\langle \mathbf{P}_j(g)(\vec{x}) + \sum_{k \leq j} \sum_{m=1}^{q-1} \mathbf{U}_k^m(g)(\vec{x}), \varphi_{j,\vec{k}}(\vec{x}) \right\rangle \quad (7.3)$$

with similar equations for all $\psi_{j,\vec{k}}^p(\vec{x})$. Since the inner products occurring in (7.3) are nothing else than the connection coefficients associated to the operator \mathbf{T} , the approximate solution to (7.1) in wavelet coordinates can be now found by solving (7.3) over all considered scales j and all possible translations \vec{k} within the data range Ω .

Recalling the results of the previous chapter, for several operators $\mathbf{T}(f)(\vec{x})$, system (7.3) becomes a system of (linear) equations given by a matrix $\tilde{\mathbf{T}}$, which is sparse in the wavelet representation. It must be now the goal to compute the GREEN's function to the problem (7.1), which is given by the inverse matrix $\tilde{\mathbf{T}}^{-1}$. Usually, this inverse matrix is a full matrix making the solution an $O(N^2)$ problem where N denotes the data size. But fortunately, the inverse operator becomes again (essentially) sparse in the wavelet domain (at least if they are of CALDERÓN-ZYGMUND or pseudo-differential type, cf. Theorem 6.9 and see also [219]) and this fact will give us efficient inversion algorithms.

7.2. Numerical Treatments — Time Discretization Schemes

This section is dedicated to study the numerical behaviour of wavelet-based GALERKIN methods applied to ill-posed problems for time discretized schemes. In the first paragraph, we will consider *preconditioning* of the (linear) systems arising from (7.3). This preconditioning has the goal to reduce the condition number of the associated matrix $\tilde{\mathbf{T}}$ in order to achieve fast convergence rates for the inversion algorithms. In the second part, we will present the different possibilities to discretize the associated operators with respect to time evolution, while the third subsection deals with the efficient implementation via dimensional reduction of the original problem. We point out that we will sometimes switch between continuous and discrete notation without further justification for keeping the presentation simpler. The reader, who is more interested in technical details is referred to the quoted literature to be calmed about these simplifications.

Usually, the numerical implementation of ill-posed problems is realized via *time discretization schemes*, that is, one introduces an evolutionary parameter

t and considers the steady-state of the diffusion system

$$\mathbf{D}_t f(\vec{x}, t) = \mathbf{T}(f)(\vec{x}, t) - g(\vec{x}) \quad (7.4)$$

instead of the problem (7.1); this corresponds to a *gradient descent* method. Discretization of the evolutionary derivative $\mathbf{D}_t f(\vec{x}, t)$ now gives

$$f(\vec{x}, t + \tau) - f(\vec{x}, t) = \tau \cdot \mathbf{T}(f)(\vec{x}, \cdot) - \tau \cdot g(\vec{x}) \quad (7.5)$$

The time discretization of the right hand side expression $\mathbf{T}(f)(\vec{x}, \cdot)$ is somewhat more versatile and shall be briefly categorized in the following. If the operator $\mathbf{T}(f)$ is linear, the straightforward choice for the discretization would be to decompose $\mathbf{T}(f)$ into

$$\mathbf{T}(f)(\vec{x}, \cdot) = \mathbf{T}_1(f)(\vec{x}, t) + \mathbf{T}_2(f)(\vec{x}, t + \tau),$$

which typically leads to classical matrix inversion algorithms like the JACOBI or the GAUSS-SEIDEL iteration method (depending on the choices of $\mathbf{T}_1(f)$ and $\mathbf{T}_2(f)$).² This issue will be discussed in the upcoming paragraph.

7.2.1. Multiscale Preconditioning of Linear Systems

For the solution of ill-posed problems, SOBOLEV spaces are the common choice as a model space, since these are well-suited to give a measure of regularity for approximate solutions on one hand but are also capable to yield descriptions for different types of noise (in the measurements). Moreover, as we shall see now, the wavelet characterization of solutions in SOBOLEV spaces will automatically lead to diagonal preconditioners for the associated linear equation systems.

Given a MRA for the function space $L_2(\mathbb{R}^n)$ with regularity s and assuming there exists some coarsest scale, say $j_0 = 0$, any function $f(\vec{x}) \in L_2(\mathbb{R}^n)$ may be written as a linear combination

$$f(\vec{x}) = \sum_{\vec{k} \in \mathbb{Z}^n} c_{\vec{k}}^0 \cdot \varphi_{0, \vec{k}}(\vec{x}) + \sum_{p=1}^{|q|-1} \sum_{j \leq 0} \sum_{\vec{k} \in \mathbb{Z}^n} d_{\vec{k}}^{j,p} \cdot \psi_{j, \vec{k}}^p(\vec{x}).$$

²Of course, the decomposition is not arbitrary. Since the iteration shall converge to a solution of the original problem (7.1), the splitting of \mathbf{T} must fulfill the *contraction condition*

$$\|(\mathbf{Id} - \tau \cdot \mathbf{T}_1)^{-1} \cdot (\mathbf{Id} - \tau \cdot \mathbf{T}_2)\| \ll 1.$$

Obviously, the convergence gets faster as this norm decreases.

Following the characterizing relation (5.17), we directly find

$$\|f(\vec{x})\|_{W_2^\sigma}^2 = |c_{0,0}|^2 + \sum_{j \leq 0} |\det \mathbf{Q}|^{-2 \cdot \sigma \cdot j} \cdot \sum_{p=1}^{|\mathbf{q}|-1} \sum_{\vec{k} \in \mathbb{Z}^n} |\langle f(\vec{x}), \psi_{j,\vec{k}}^p(\vec{x}) \rangle|^2. \quad (7.6)$$

for $|\sigma| \leq s$. Since the squared inner products between the function $f(\vec{x})$ and the wavelets correspond to the l_2 -norm of the wavelet coefficients, we have the norm equivalence

$$\|f(\vec{x})\|_{W_2^\sigma}^2 = |c_{0,0}|^2 + \|\Pi_\sigma \cdot \vec{d}_{j,\vec{k}}\|_{l_2}^2, \quad (7.7)$$

where $\vec{d}_{j,\vec{k}}$ denotes the vector of expansion coefficients and Π_σ is a diagonal matrix with entries $|\det \mathbf{Q}|^{-\sigma \cdot j}$. This fact can be understood in the sense that the rescaled functions $|\det \mathbf{Q}|^{\sigma \cdot j} \cdot \psi_{j,\vec{k}}^p(\vec{x})$ constitute a wavelet frame for the SOBOLEV space $W_2^\sigma(\mathbb{R}^n)$ as long as $|\sigma| \leq s$.

Recalling now our original problem (7.3), we have to find the solution to a linear system

$$\tilde{\mathbf{T}} \cdot \vec{f}_{j,\vec{k}} = \vec{g}_{j,\vec{k}}, \quad (7.8)$$

or its time-discretized variant (7.5), where $\vec{f}_{j,\vec{k}}$ and $\vec{g}_{j,\vec{k}}$ are the wavelet expansion coefficient vectors of the unknown $f(\vec{x})$ and the measurement $g(\vec{x})$ and $\tilde{\mathbf{T}}$ is the matrix representation of the involved operator \mathbf{T} . Since the matrix $\tilde{\mathbf{T}}$ is assumed to be rather huge, iterative methods like GAUSS-SEIDEL iterations or (over-)relaxation iteration depending on the discretization chosen in (7.5) and on the splitting of $\tilde{\mathbf{T}}$ are required. Now, it is a well-known fact from numerical analysis, that the convergence rate of such techniques is highly depending on the *condition number* $\kappa(\tilde{\mathbf{T}})$ which is given by the absolute value of the quotient between the largest and the smallest eigenvalue of $\tilde{\mathbf{T}}$, or equivalently by

$$\kappa(\tilde{\mathbf{T}}) = \|\tilde{\mathbf{T}}\| \cdot \|\tilde{\mathbf{T}}^{-1}\|.$$

Unfortunately, the condition number $\kappa(\tilde{\mathbf{T}})$ can be shown to behave (usually) like

$$\kappa(\tilde{\mathbf{T}}) \sim |\det \mathbf{Q}|^{-2 \cdot c \cdot \sigma \cdot j}$$

for a j -scale matrix representation of \mathbf{T} , where the (positive) constant c depends on the kind of operator [56]. This means, that for all $c > 0$, the convergence rate of the iteration decreases exponentially the finer the scale j is chosen.³ However,

³The reader should take care that this problem is not overcome by taking a scale equal or close to zero since this gives rather non-sparse matrix representations of the inverse operators. High sparsity for the matrices requires a finer scale resolution.

if the operator \mathbf{T} behaves as an isometry between the spaces W_2^σ and its dual $W_2^{-\sigma}$, that is

$$\|f(\vec{x})\|_{W_2^\sigma}^2 \sim \|\mathbf{T}f(\vec{x})\|_{W_2^{-\sigma}}^2 \quad (7.9)$$

for all $f(\vec{x}) \in W_2^\sigma$, then the matrix $\Pi_\sigma^{-1} \cdot \tilde{\mathbf{T}} \cdot \Pi_\sigma^{-1}$ representing the operator \mathbf{T} on the space W_2^σ is uniformly bounded since it acts as an isometry. In other words,

$$\kappa(\Pi_\sigma^{-1} \cdot \tilde{\mathbf{T}} \cdot \Pi_\sigma^{-1}) \sim O(1), \quad (7.10)$$

which gives rise to a uniform convergence rate indepently of the chosen scale j for the *preconditioned system*

$$\Pi_\sigma^{-1} \cdot \tilde{\mathbf{T}} \cdot \Pi_\sigma^{-1} \cdot \vec{e}_{j,\vec{k}} = \Pi_\sigma^{-1} \cdot \vec{g}_{j,\vec{k}}, \quad (7.11)$$

which solves the problem (7.8) in the SOBOLEV space of smoothness σ .⁴ Hereby, $\vec{e}_{j,\vec{k}}$ denotes the SOBOLEV space representation of $\vec{f}_{j,\vec{k}}$, which is given by

$$\vec{e}_{j,\vec{k}} = \Pi_\sigma \cdot \vec{f}_{j,\vec{k}}.$$

It remains the task to find the right SOBOLEV exponent σ such that the isometry condition (7.9) is indeed satisfied. This can be done under application of (B.11) to the wavelet representation (5.5). Inserting $p = q = 2$ and the (in practice finite scale) representations of $f(\vec{x})$ and $\mathbf{T}(f)(\vec{x})$ into (B.11), one directly gets

$$\|f(\vec{x})\|_{W_2^\sigma}^2 = \sum_{j \in \mathbb{Z}} |\det \mathbf{Q}|^{2 \cdot j \cdot \sigma} \cdot \sum_{\vec{k}} \sum_{p=1}^{|q|-1} |d_{\vec{k}}^{j,p}|^2$$

and using the theory of the previous chapter one also obtains

$$\|\mathbf{T}(f)(\vec{x})\|_{W_2^{-\sigma}}^2 = \sum_{j \in \mathbb{Z}} |\det \mathbf{Q}|^{-2 \cdot j \cdot \sigma} \cdot \sum_{i \in \mathbb{Z}} \sum_{\vec{k}} \sum_{\vec{l}} \sum_{p=1}^{|q|-1} \sum_{r=1}^{|q|-1} |d_{\vec{k}}^{j,p}|^2 \cdot |t_p^r(j, \vec{k}, i, \vec{l})|^2,$$

where $t_p^r(\cdot, \vec{\cdot}, \cdot, \vec{\cdot})$ denotes the connection coefficients associated to the considered operator \mathbf{T} . Equating these expressions yields

$$|\det \mathbf{Q}|^{4 \cdot j \cdot \sigma} = \sum_{\vec{i} \in \mathbb{Z}} \sum_{\vec{m}} \sum_{r=1}^{|q|-1} |t_p^r(0, \vec{0}, \vec{i}, \vec{m})|^2 \quad \forall p \in \{1, \dots, |q|-1\}, \quad (7.12)$$

⁴One automatically obtains $f(\vec{x}) \in W_2^\sigma(\mathbb{R}^n)$ and $g(\vec{x}) \in W_2^{-\sigma}(\mathbb{R}^n)$. For positive σ , this especially means that we assured that the solution $f(\vec{x})$ possesses some regularity, while on the other hand, the measurements $g(\vec{x})$ are allowed to have negative regularity. Therefore, this setting is even capable to model noise within the measurements.

which gives the correct value for σ by simple calculations. One should note here, that due to the symmetry properties of connection coefficients, equation (7.12) can be solved for all p simultaneously. To give an example, it is well-known that $2m$ -th order elliptic boundary value problems require $\sigma = m$ [56].

We conclude that preconditioning is a valuable tool to reduce the computational amount for calculating the GREEN's function for problems of the kind (7.1); regarding computer vision applications, this is of special interest for linear and non-linear diffusion processes occurring in *scale space* representations (see next chapter). However, the huge operator matrix sizes (for example, an image of size 256×256 requires operator matrices of size 65536×65536 in the standard form and of 87380×87380 in the non-standard form) might be prohibitive in certain hardware surroundings and make these methods not the optimal choice, especially for more intricate problems as they are posed in e.g. optical flow estimations.

7.2.2. Time Discretization of (Possibly) Nonlinear Operators

The preconditioning method has lead to convergent iterative inversion algorithms relying on uniformly bounded condition numbers. But we also had to put up with huge storage requirements and with the fact that this only works with linear(ized) equation systems. Consequently, we will develop alternatives to circumvent these drawbacks.

In all of the following considerations, we will assume that the operator $\mathbf{T}(f)(\vec{x}, \cdot)$ is not necessarily linear but has at least a *linear factor*, that is, we may rewrite $\mathbf{T}(f)(\vec{x}, \cdot)$ as

$$\mathbf{T}(f)(\vec{x}, \cdot) = \mathbf{N}(f)(\vec{x}, \cdot) \cdot \mathbf{L}(f)(\vec{x}, \cdot), \quad (7.13)$$

where $\mathbf{L}(f)(\vec{x}, \cdot)$ is a linear operator.⁵ Time-discretization of the right-hand-side of the system (7.5) can now be done in *explicit*, *semi-implicit* and *fully implicit* manner. Since the latter case would lead to a (usually very expensive) inversion of a nonlinear operator if $\mathbf{T}(f)(\vec{x}, \cdot)$ is nonlinear, it is not considered here. Moreover, it will turn out that even semi-implicit schemes fulfill all the stability and convergence requirements that are necessary for our applications. Explicit discretization in (7.5) would lead to an iteration scheme

$$f(\vec{x}, t + \tau) = (\mathbf{Id} + \tau \cdot \mathbf{T})(f)(\vec{x}, t) - \tau \cdot g(\vec{x}), \quad (7.14)$$

⁵We point out that this decomposition must not necessarily be done by such a multiplicative splitting. The important thing is just the possibility to split off a linear part in order to arrive at a linear system of equations that describe a time-discretized version of the original process.

which does not require matrix inversions and might look attractive at a first glance. But one easily sees, that if τ becomes larger, the matrix norm of the operator $(\mathbf{Id} + \tau \cdot \mathbf{T})(f)$ may exceed the radius of convergence given by the critical value 1, which leads to instability in the numerical process. This fact forces very small iteration steps τ , in general leading to rather slow convergence; we will illustrate this phenomenon in more detail in the next chapter. Anyway, for the applications we have in mind, we use explicit discretizations only in very few special situations, our method of choice will be primarily semi-implicit discretization.

Discretizations of this type rely on the operator decomposition (7.13) by associating the evaluated step at time t to the nonlinear part $\mathbf{N}(f)(\vec{x}, t)$ and the *prediction* at time $t + \tau$ to the linear part $\mathbf{L}(f)(\vec{x}, t + \tau)$. Semi-implicit discretization schemes are (usually) much less sensitive against the size of the evolution step τ , which makes them attractive from the computational point of view; they are given by

$$f(\vec{x}, t + \tau) = (\mathbf{Id} - \tau \cdot \mathbf{N}(f)(\vec{x}, t) \cdot \mathbf{L})^{-1} (f(\vec{x}, t) - \tau \cdot g(\vec{x})). \quad (7.15)$$

The major work to be done is to invert the operator $\mathbf{Id} - \tau \cdot \mathbf{N}(f)(\vec{x}, t) \cdot \mathbf{L}$ in each time step. This may seem even more expensive than doing many small explicit steps, but fortunately, the special sparse structure the operators have in wavelet coordinates gives rise to very efficient inversion algorithms. One such example is the preconditioning method demonstrated in the previous subsection, another even more attractive method will be presented now.

7.2.3. Additive Operator Splitting

Suppose, we are given a scheme as in (7.15). Especially in higher dimensions, the associated operator matrices $\mathbf{Id} - \tau \cdot \mathbf{N}(f)(\vec{x}, t) \cdot \mathbf{L}$ become prohibitively large in this situation (as mentioned before); moreover, due to the coercive reordering of a multidimensional data set into a vector, there occur far off-diagonal elements in the operator matrices, which prevent efficient direct inversions. For this reason, it would be nice to replace the multidimensional operator by a sum of (intrinsically) onedimensional operators that describe the same diffusion process. In other words, a decomposition

$$(\mathbf{N}(f)(\vec{x}, t) \cdot \mathbf{L})(f)(\vec{x}, \cdot) = \sum_{j=1}^m \mathbf{N}_j(f_j)(\vec{x}, \cdot) \quad (7.16)$$

is desired. Hereby, $f_j(\vec{x}) \in \mathbb{R}^{N_{j(m)}}$ denotes a set of *subvectors* of the discretized image data $f(\vec{x}) \in \mathbb{R}^{N_1 \cdot N_2 \cdots N_n}$.⁶ If an *additive operator splitting* as in (7.16) is known, then one may consider the diffusion process

$$f(\vec{x}, t + \tau) = \frac{1}{m} \cdot \sum_{j=1}^m (\mathbf{Id} - m \cdot \tau \cdot \mathbf{N}_j(f_j)(\vec{x}, t) \cdot \mathbf{L})^{-1} (f_j(\vec{x}, t) - \tau \cdot g_j(\vec{x})) \quad (7.17)$$

instead of (7.15). It can be shown under usage of a TAYLOR series, that the schemes (7.15) and (7.17) have the same approximation order in the spatial and the evolutionary variable [236] [237] [176]. This means that both schemes are well-suited approximations of the original problem (7.1). But while (7.15) requires the inversion of a matrix of size $(N_1 \cdot N_2 \cdots N_n) \times (N_1 \cdot N_2 \cdots N_n)$ containing far off-diagonal elements in each step, the additive operator splitting scheme only requires the inversion of m narrowbanded matrices of the size $N_{j(m)} \times N_{j(m)}$, this can be done very efficiently, even in linear time. (Usually, the value $N_{j(m)}$ is equal to N_j , but for onedimensional discretizations along certain diagonals, the data vector is shortened due to the rectangular structure of the discretion grid giving $N_{j(m)} \leq N_j$, see also the example below.)

EXAMPLE* 8 For a better understanding, we will illustrate the additive splitting in a 2D example with a 3×3 operator mask now. Assume therefore, that the discretized operator stencil at the grid point $[i, j]$ shall be given by

$$\mathbf{N}_{i,j} = \begin{bmatrix} n_{i-1,j-1} & n_{i,j-1} & n_{i+1,j-1} \\ n_{i-1,j} & n_{i,j} & n_{i+1,j} \\ n_{i-1,j+1} & n_{i,j+1} & n_{i+1,j+1} \end{bmatrix}$$

and zeros elsewhere. This representation may be split into the sum

$$\begin{aligned} \mathbf{N}_{i,j} = & \begin{bmatrix} n_{i-1,j-1} & 0 & 0 \\ 0 & \frac{n_{i,j}}{4} & 0 \\ 0 & 0 & n_{i+1,j+1} \end{bmatrix} + \begin{bmatrix} 0 & 0 & n_{i+1,j-1} \\ 0 & \frac{n_{i,j}}{4} & 0 \\ n_{i-1,j+1} & 0 & 0 \end{bmatrix} \\ & + \begin{bmatrix} 0 & n_{i,j-1} & 0 \\ 0 & \frac{n_{i,j}}{4} & 0 \\ 0 & n_{i,j+1} & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ n_{i-1,j} & \frac{n_{i,j}}{4} & n_{i+1,j} \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

⁶In multigrid algorithms, multiplicative splittings of operators also play an important role. However, since the associated matrices do not commute in general, these are not invariant under affine transformations and are thus less suited in image processing. Consequently, we will restrict ourselves to additive operator splittings in this work.

which are intrinsically onedimensional operators acting on the horizontal, vertical and diagonal parts of the data structure. These are represented by a sum of tridiagonal matrices multiplied by subvectors f_j of the data matrix f in terms of (7.17) and give rise to the announced simplified inversion. In order to gain even more computational efficiency, the single operators may be inverted in parallel supposed a corresponding hardware is available. Similar operator splittings are also possible for larger stencils in a straightforward manner giving narrowbanded onedimensional operators; some care has to be taken on diagonals of degree $\alpha \neq k \cdot \pi/4$ with $k \in \mathbb{Z}$, since these give shorter stencils and shorter data subvectors f_j .

One directly notes that additive operator splitting is only suited for single-scale represented operators, since it relies on nearly diagonal operators. For a multiscale representation, one obtains far off-diagonal elements corresponding to scale interactions (standard form) or to lowpass-highpass and highpass-highpass interactions (both forms). This makes additive operator splitting as it is defined above unsuited on wavelet representations. Consequently, our next goal is to combine the computational power of additive operator schemes with that of the wavelet framework. To do this, we will substitute the linear prediction part $\mathbf{L}(f)(\vec{x}, t)$ in (7.17) by a prediction and an updating part yielding

$$\mathbf{L}(f)(\vec{x}, \cdot) = \mathbf{L}_1(f)(\vec{x}, t) + \mathbf{L}_2(f)(\vec{x}, t + \tau). \quad (7.18)$$

Application of additive operator splitting then gives the iterative scheme

$$\begin{aligned} f(\vec{x}, t + \tau) &= \frac{1}{m} \cdot \sum_{j=1}^m (\mathbf{Id} - m \cdot \tau \cdot \mathbf{N}_j(f_j)(\vec{x}, t) \cdot \mathbf{L}_2)^{-1} \cdot \\ &\quad \left((\mathbf{Id} - \tau \cdot \mathbf{N}(f)(\vec{x}, t) \cdot \mathbf{L}_1)(f_j)(\vec{x}, t) - \tau \cdot g_j(\vec{x}) \right). \end{aligned} \quad (7.19)$$

Recalling the standard and non-standard representations (6.21) and (6.22), we may use (7.18) to split the wavelet form of the operator $\mathbf{L}(f)$ into a near-diagonal and a far off-diagonal part. For the standard form, one chooses

$$\begin{aligned} \mathbf{L}_1(f) &= \sum_{i=1}^j \sum_{k=1, k \neq i}^j \sum_{m=1}^{|q|-1} \sum_{p=1}^{|q|-1} \mathbf{U}_k^p \mathbf{L} \mathbf{U}_i^m(f) \quad \text{and} \\ \mathbf{L}_2(f) &= \mathbf{P}_j \mathbf{L} \mathbf{P}_j(f) + \sum_{i=1}^j \sum_{m=1}^{|q|-1} \sum_{p=1}^{|q|-1} \mathbf{U}_i^p \mathbf{L} \mathbf{U}_i^m(f). \end{aligned}$$

For the non-standard form, the choices

$$\begin{aligned} \mathbf{L}_1(f) &= \sum_{i=1}^j \sum_{p=1}^{|q|-1} \left(\mathbf{U}_i^p \mathbf{L} \mathbf{P}_i(f) + \mathbf{P}_i \mathbf{L} \mathbf{U}_i^p(f) \right) \quad \text{and} \\ \mathbf{L}_2(f) &= \mathbf{P}_j \mathbf{L} \mathbf{P}_j(f) + \sum_{i=1}^j \sum_{m=1}^{|q|-1} \sum_{p=1}^{|q|-1} \mathbf{U}_i^p \mathbf{L} \mathbf{U}_i^m(f) \end{aligned}$$

are appropriate. Hereby, we used the notation of Chapter 6. With this decompositions, the extended additive operator splitting iteration scheme (7.19) can be applied directly in wavelet coordinates in both representation forms. Anyway, under usage of regularized nonlinear scale space embeddings, an even more efficient wavelet based approximative variant of the additive operator splitting method can be found. This extension will be developed in the following chapter.

7.3. Related Work

The idea of using wavelets to solve operator equations goes back to the fundamental paper [20], much work in this direction has also been done by RIEDER, WELLS and ZHOU (see e.g. [188]). Preconditioning of the corresponding equation systems in the wavelet domain originates in the work of DAHMEN and KUNOTH [130] [57], similar ideas to these are shared by *multigrid methods*, another efficient tool for inversion of large (sparse) matrices [103].

The additive operator splitting is a result of optimizing the performance of nonlinear diffusion processes in image processing, it is mainly due to WEICKERT [237] [236]. The contribution of this work consists in taking a more *unified view* by combining the advantages of wavelet representations of operators with the idea of operator splitting to solve ill-posed problems.

Chapter Summary

This short chapter gave an introduction to the solution of inverse problems with wavelet methods. The basic idea was to consider the given original problem not globally but only on the local wavelet spaces by the so-called GALERKIN projection. This method turned out to have a close connection with the operator representations introduced earlier in this thesis and has led us to the problem of efficient inversions. Afterwards, we have shown the possibility of preconditioning the given linear(ized) systems leading to smaller condition numbers and thus to faster iterative algorithms. Another possibility we have investigated was the additive operator splitting method that made use of a dimensional decomposition of the originally multidimensional operator into a sum of intrinsically onedimensional operators. By changing the order of summing and inverting (which is only approximately true), the whole procedure became a linear time

inversion algorithm. At the end of this chapter, we have sketched a possibility to combine the advantages of additive operator splitting with the wavelet framework.

III

MULTIDIMENSIONAL WAVELETS
IN COMPUTER VISION

Chapter 8

LINEAR AND NONLINEAR SCALE SPACES

It is sometimes said that the great discovery of the nineteenth century was that the equations of nature are linear, and the great discovery of the twentieth century is that they are not.

—THOMAS W. KÖRNER [125]

It is an inherent property of any real-world object that it is a meaningful entity only over a finite range of resolution levels. This physical fact does directly carry over to *objects* in images, it builds the fundamental idea of the *scale space* concept: The physical observation of an image requires the usage of a measuring device (typically a camera) with a certain (usually adjustable) aperture. In this situation, one is confronted with the tradeoff problem that if one is interested in focussing on small details, the aperture must be chosen accordingly narrow which means that less of the whole image will be registered. On the other hand, the usage of a larger aperture receives bigger parts of the whole image but comes along with the drawback of coarsening the details. Generally, the observer has no a priori knowledge about the right choice of scale to make a desired observation. Consequently, one should embed the original image into a one-parameter family of images representing the original one on various scales. Under certain axiomatic conditions which will be discussed in the following, such a one-parameter family is called a *scale space representation* of an image.

This chapter is organized as follows. In the first section, we will introduce the axiomatic setting leading to linear scale space representations. We will introduce the heat conduction equation as a basic model for scale space theories. Afterwards, some of the axioms are slightly *weakened* leading to linear wavelet scale spaces, which have a specially simple structure. The second part of this chapter is dedicated to nonlinear scale spaces. In particular, we will discuss

the PERONA-MALIK approach and a more differential geometric point of view leading to anisotropic diffusion processes. Again, a particular simple definition of nonlinear scale spaces can be found by wavelet thresholding the image data with respect to BESOV norms.

8.1. Linear Scale Spaces

As stated in the introduction of this chapter, the scale space representation shall be an embedding of the original image into a whole family of images containing rescaled versions of the input. Hereby, the increasing of scale has to be interpreted as a simplification of the original image by removing *disturbing* or *unnecessary* details. But on the other hand, one should also assure that this process does not create new structures, which obviously contradicts the physical aspect of the model and would additionally not be desirable in any kind of simplification. We give the following definition.

DEFINITION 8.1 (LINEAR SCALE SPACE [138]) *Let $f(\vec{x})$ with $\vec{x} \in \mathbb{R}^n$ be a given image and $\{U_t(\vec{x}) \mid t \geq 0\}$ a one-parameter family of images such that $U_0(\vec{x}) = f(\vec{x})$. Then, $U_t(\vec{x})$ is called a scale space representation of the image $f(\vec{x})$ if the following conditions are satisfied.*

i.) *The semi-group property*

$$U_{s+t}(\cdot) = U_s(\cdot) * U_t(\cdot) \quad s, t \in \mathbb{R}_{\geq 0}.$$

ii.) *The symmetry property*

$$U_t(\vec{x}) = U_t(S_n(\vec{x})) \quad \forall t \geq 0, \vec{x} \in \mathbb{R}^n,$$

if $S_n(\vec{x})$ is any sign or component permutation of \vec{x} .

iii.) *The continuity requirement*

$$\lim_{t \rightarrow 0} \|U_t(\vec{x}) - f(\vec{x})\| = 0.$$

iv.) *The nonenhancement properties*

$$\begin{aligned} D_t U_{t_0}(\vec{x}) &\leq 0 && \text{if } U_{t_0}(\vec{x}) \text{ is a local maximum and} \\ D_t U_{t_1}(\vec{x}) &\geq 0 && \text{if } U_{t_1}(\vec{x}) \text{ is a local minimum.} \end{aligned}$$

This last condition guarantees that indeed no new structures are created as the scale parameter t increases.

Remark. The notion of scale space was originally introduced to the computer vision community in the fundamental works [244] and [124]. However, there

exist various differing axiomatics to define linear scale space, see e.g. [236] for an overview. All definitions lead to the same fundamental solution, namely a normalized GAUSSian function of standard deviation $\sqrt{2t}$ which is easily derived. In other words, the convolution

$$U_t(\vec{x}) = \int_{\mathbb{R}^n} f(\vec{y}) \cdot T_t(\vec{x} - \vec{y}) \, d\vec{y} \quad (8.1)$$

with

$$T_t(\vec{x}) = \frac{1}{2 \cdot \pi \cdot t^2} \cdot e^{-\frac{|\vec{x}|^2}{2 \cdot t^2}}$$

defines the linear scale space representation of $f(\vec{x})$. This solution motivates the commonly used term *GAUSSian scale space* for a linear scale space. It is evident that equation (8.1) describes a smoothing process, moreover $U_t(\vec{x}) \in \mathcal{C}^\infty(\mathbb{R}^n)$ holds true for all $t > 0$.

8.1.1. The Heat Conduction Equation and Gaussian Scale Space

From a physical point of view, an image smoothing process may be viewed as diffusion of the energy distribution within a closed system over a period of time. Such a diffusion process obeys the *heat conduction equation*

$$\mathbf{D}_t U_t(\vec{x}) = \operatorname{div}_{\vec{x}}(g(\vec{x}) \cdot \nabla_{\vec{x}} U_t(\vec{x})) , \quad (8.2)$$

where the function $g(\vec{x})$ describes the specific heat conductivity at a point \vec{x} . Choosing homogeneous diffusion over the whole image range, that is $g(\vec{x}) \equiv 1$, equation (8.2) reduces to the homogeneous heat conduction equation

$$\mathbf{D}_t U_t(\vec{x}) = \Delta_{\vec{x}} U_t(\vec{x}) . \quad (8.3)$$

It is a well-known result from the theory of PDEs that the problem (8.3) together with the *boundary condition*

$$U_0(\vec{x}) = f(\vec{x})$$

possesses (up to a scaling) the unique fundamental solution given by (8.1), provided $f(\vec{x})$ satisfies a moderate asymptotic condition. The physical interpretation of the result is as follows: It is formulated in the first fundamental theorem of thermodynamics that the energy within a closed system remains constant over time, which is reflected by the fact that the fundamental solution is normalized. The second fundamental theorem of thermodynamics postulates non-reduction of entropy in any heat distributing process. This can be seen in parallel to the nonenhancement-of-structure property in the definition of linear scale space. Finally, the third fundamental theorem of thermodynamics, the NERNST theorem, says that any heat diffusion process runs into a constant steady state.

Indeed, one easily verifies the identity $\lim_{t \rightarrow \infty} \mathcal{U}_t(\vec{x}) = \text{const.}$

GAUSSian smoothing is applied in many situations in image processing. For example, the commutativity with derivative operators

$$\mathbf{D}_{\vec{x}}^{\vec{m}}(\mathcal{T}_t * f)(\vec{x}) = (\mathbf{D}_{\vec{x}}^{\vec{m}}(\mathcal{T}_t)(\vec{x})) * f(\vec{x}) \quad , \quad \vec{m} \in \mathbb{N}_0^n ,$$

makes the (usually) ill-posed differentiation of images well-posed in the scale space embedding, since the convolution function $\mathcal{T}_t(\vec{x})$ is a \mathcal{C}^∞ -function. This means that substituting derivatives by these GAUSSian derivatives causes strong regularization. We will see in the next section, how this property will be utilized to turn ill-posed nonlinear scale space characterizations into well-posed ones. Other applications of GAUSSian smoothing in image processing are e.g. edge detection via LAPLACians of GAUSSians (also known as the MARR-HILDRETH operator [152]), multiscale segmentation or the problem of finding the relevant scales by entropy analysis [211]. These issues shall not be discussed here in further detail.

We close this paragraph by some remarks concerning the numerical implementation of GAUSSian smoothing. There are two main possibilities. Following [138] one uses a discretized version of the convolution in (8.1). Hereby, the (principally infinite) GAUSSian filter has to be approximated in a well-suited way; several suggestions how this can be done are found in the monograph [138]. Since the calculation of a discrete convolution requires $O(n)$ critical operations, this method gives a linear time approximation of GAUSSian smoothing. The second possibility consists in a time-discretization of the operator equation (8.3). This can be implemented efficiently using e.g. finite differences or wavelet representations of operators and optimized under usage of additive operator splittings also leading to a linear time algorithm [237]. Consequently, the kind of implementation is merely a question of taste.

8.1.2. Linear Wavelet Scale Space

We have seen that the scale space axiomatic given by Definition 8.1 leads to the GAUSSian as the unique solution. Since the convolution of a signal (image) with a GAUSSian corresponds to a lowpass filtering, one may ask whether a larger class of lowpass convolution filters is accessible if certain requirements of Definition 8.1 are relaxed without destroying the physical interpretation given above in terms of homogeneous heat distribution. Clearly, the symmetry property and the nonenhancement requirement are necessary to stay consistent with the physical model and shall be kept. But since digital images are discrete entities, it is not a model violation to reduce the semi-group property to discrete multiscale parameters. In other words, the semi-group requirement i.) in

Definition 8.1 may be replaced by a *discrete multiscale* property

$$\mathcal{U}_s(\cdot) = f(\cdot) * \varphi(\mathbf{Q}^s \cdot), \quad s \in \mathbb{N}_0$$

with a scaling matrix \mathbf{Q} , a discrete scaling parameter s and a suited smoothing function $\varphi(\cdot)$ (see Lemma 8.2 below).

In this situation, the continuity requirement iii.) can obviously be completely dropped. Under this relaxation, it is easy to show the following result.

LEMMA* 8.2 *Suppose $[\mathbf{h}_{\vec{k}}]_{\vec{k} \in \mathbb{Z}^n}$ is a symmetric, nonnegative and unimodal lowpass scaling filter with an associated scaling matrix \mathbf{Q} and a scaling function $\varphi(\vec{x})$. Assume further that the considered image $f(\vec{x})$ lies in some approximation space of the related MRA, say $f \in V_0$ (without loss of generality).¹ Recalling the definition of the projection operator $\mathbf{P}_j(f)$ in Chapter 5, the assignment*

$$\mathcal{U}_j(\vec{x}) = \mathbf{P}_0(\mathbf{P}_j(f))(\vec{x}), \quad j \in \mathbb{N}_0, \quad (8.4)$$

*induces a discrete linear wavelet scale space in the relaxed sense.*²

Remark. A certain special case of this definition is the scale space derived from B-splines as developed in [230]. Due to the discrete character of the scaling parameter j , we can not give a direct formulation of linear wavelet scale space in terms of a PDE. However, since it stems from a symmetric, nonnegative filter, a linear wavelet scale space may be interpreted as a discrete approximation of a TIKHONOV regularization process [220] if a *reaction term* like $(\mathcal{U}_t(\vec{x}) - f(\vec{x}))$ would be added. The class of TIKHONOV regularization processes is physically described by elliptic even PDEs of the type

$$\mathbf{D}_t \mathcal{U}_t(\vec{x}) = (\mathcal{U}_t(\vec{x}) - f(\vec{x})) + \sum_{\vec{k} \in \mathbb{N}^n} \alpha_{\vec{k}} \cdot (\mathbf{D}_{\vec{x}}^{\vec{k}} \mathcal{U}_t(\vec{x}))^2, \quad \alpha_{\vec{k}} \geq 0.$$

Proof of Lemma 8.2. The requirements for symmetry, nonnegativity and unimodality are well-known to be coercive for all scale space representations [236], they are necessary to guarantee nonenhancement of structures and stability. Together with the fact that the wavelet transform preserves the energy norm of a signal, this gives sufficiency to assure nonenhancement and stability. The new multiscale property i'.) is satisfied by construction which concludes

¹From the continuous point of view, this is a restrictive condition. However, since one usually deals with discrete images, the embedding $f \in V_0$ is quite natural, see also the approximation results in Chapter 5.

²The backprojection \mathbf{P}_0 into the space V_0 guarantees that all $\mathcal{U}_j(\vec{x})$ are of the same size in the case of discrete data. If this backprojection would not be done, one ends up with a *lowpass pyramid* representation as it was proposed by BURT in [28]. This multiscale embedding has the drawback of reducing the information content with increasing scale j .

the proof. \square .

It is clear that if $\varphi(\vec{x}) \in \mathcal{C}^s$ holds, the same is true for all $\mathcal{U}_j(\vec{x})$ with $j > 0$. Consequently, linear wavelet scale space representations may also be utilized to regularize differentiation supposed the smoothness index s is large enough. Although the definitions of GAUSSian and linear wavelet scale space seem to be quite different at a first glance, their results look rather similar in practice, see Figure 8.1.

Of course, GAUSSian scale space possesses the advantage to be accessible for every desired scale parameter $t \in \mathbb{R}_+$, while the linear wavelet scale space is limited to integer dilations. Nevertheless, its close connection with standard wavelet representations including efficient implementations of operators and operator equations makes it an attractive choice anyway.

We shall finally mention that there exist further different approaches to linear scale spaces by replacing the semi-group property by another scale-oriented requirement. For example, it was recently shown that a linear scale space can be derived from an inverse wave propagation process via the LAPLACE equation [84]. Due to its inverse character, this construction is difficult to handle numerically, but on the other hand it exhibits some nice parallels to the mathematical theory of RIESZ transforms and CALDERÓN-ZYGMUND operators.

8.2. Nonlinear Scale Space Extensions

Beings low in the scale of nature are more variable than those which are higher.

—CHARLES DARWIN [59]

The example of Figure 8.1 demonstrated that a linear scale space embedding smoothens out details with high frequencies from images (as desired). Unfortunately, it also blurred details on larger scales, since the process can not *differ* between high frequencies occurring from small scale features and high frequencies belonging to e.g. step edges of larger scale features. Another phenomenon that happens to appear is the dislocation of structures under blurring as it was demonstrated in [244]. Consequently, an adaption of the smoothing process is strongly indicated.

8.2.1. The Perona-Malik Approach

The first to overcome the problem of uniform smoothing were PERONA and MALIK [180]. They originally proposed to consider the diffusion as an *inhomogeneous* heat distribution process by replacing the constant specific heat



Figure 8.1. Comparison between linear scale spaces. *Left.* GAUSSIAN scale space with scaling parameter $t = 0, 2, 4, 8$. *Right.* Linear wavelet scale space with $j = 0, 1, 2, 3$. Scaling function given by cubic box spline with smoothness exponent $s = 2$. Note how the strong regularization property compensates the graininess of the photo paper already after one step.

conductivity $g(\vec{x})$ in (8.2) by

$$g(\vec{x}) = g(|\nabla f(\vec{x})|) = \frac{1}{1 + |\nabla f(\vec{x})|^2 / \lambda^2}, \quad \lambda > 0. \quad (8.5)$$

Since this specific heat conductivity is rapidly decreasing with increasing magnitude of the image gradient, the smoothing effect is reduced the larger the gradient of the signal becomes (edge preservation). The additional variable λ is introduced to adjust the sensitivity of the diffusion against $|\nabla f(\vec{x})|$, it is usually called *contrast parameter*. The behaviour of nonlinear diffusion under the conductivity (8.5) has been thoroughly studied, e.g. in [181]. Investigating the monotony behaviour of the *flux function*

$$\phi(|\nabla f(\vec{x})|) = |\nabla f(\vec{x})| \cdot g(|\nabla f(\vec{x})|)$$

and introducing *gauge coordinates*, the heat inhomogeneous conduction equation (8.2) may be rewritten as

$$\mathbf{D}_t \mathbf{U}_t(\vec{x}) = \phi'(|\nabla \mathbf{U}_t(\vec{x})|) \cdot \mathbf{D}_{\xi_1}^2 \mathbf{U}_t(\vec{x}) + g(|\nabla \mathbf{U}_t(\vec{x})|) \cdot \left(\sum_{j=2}^n \mathbf{D}_{\xi_j}^2 \mathbf{U}_t(\vec{x}) \right), \quad (8.6)$$

where the first coordinate ξ_1 is determined by the flowlines of $\mathbf{U}_t(\vec{x})$. It is given in normalized form by

$$\xi_1 = \frac{\langle \nabla \mathbf{U}_t(\vec{x}), \vec{x} \rangle}{|\nabla \mathbf{U}_t(\vec{x})|}.$$

The choice of the *normal hyperplane*, i.e. the span of the *generalized isophotes* ξ_2, \dots, ξ_n is determined by the orthogonal complement of the image gradient.³ Considering the course of the PERONA-MALIK flux in Figure 8.2, it becomes evident that the PERONA-MALIK process gives forward diffusion in the normal hyperplane and for $|\nabla f(\vec{x})| \leq \lambda$ also along the flowlines of maximum variation. Backward diffusion along flowlines occurs for $|\nabla f(\vec{x})| > \lambda$. Although it is a nice property in the context of image enhancement, this forward-backward flow raises theoretical problems, since inverse heat diffusion processes are well-known to be unconditionally instable [120]. Till today, it is not known, whether the PERONA-MALIK diffusion generally possesses a solution even in a weak sense and additionally, it can be shown that a solution (assumed one exists) must not be unique. In the following, we will quote some possibilities how to regularize nonlinear heat diffusion.

One possibility to regularize nonlinear diffusion is to make another choice for the conductivity function $g(\cdot)$ such that $\phi(\cdot)$ becomes monotonously increasing. One such example is the regularizer stemming from a convex variational formulation as studied by SCHNÖRR [201]. In terms of a specific heat conductivity,

³This generalizes the twodimensional representation of the PERONA-MALIK diffusion as given in [3] to higher dimensions. Note that this generalization leaves $n - 2$ degrees of freedom in dimensions n larger than two. These can be reduced by assigning some differential geometric interpretation (curvature, torsion etc.) to the spanning vectors of the normal hyperplane.

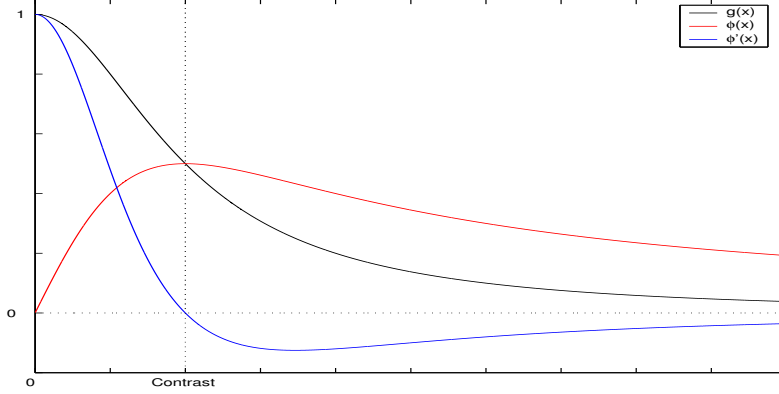


Figure 8.2. PERONA-MALIK specific diffusivity $g(\cdot)$, flux $\phi(\cdot)$ and first flux derivative $\phi'(\cdot)$.

it is defined by

$$g(|\nabla f(\vec{x})|) = \begin{cases} 1, & |\nabla f(\vec{x})| \leq \lambda \\ \frac{\lambda \cdot (1 - c) + c \cdot |\nabla f(\vec{x})|}{|\nabla f(\vec{x})|}, & |\nabla f(\vec{x})| > \lambda \end{cases} \quad (8.7)$$

with a constant $1 > c > 0$ and the contrast parameter $\lambda > 0$. Due to the convexity of the considered functional, it immediately follows that nonlinear diffusion with the conductivity (8.7) always possesses a unique global solution. Its drawbacks are on one hand that the diffusion is not continuous due to the jump in the flux derivation $\phi'(\cdot)$ as shown in Figure 8.3 and on the other hand, the positivity of $\phi'(\cdot)$ leads to forward diffusion for all image gradients, i.e. over a long period of time, all edges are removed and one ends up with a constant steady state.

Another idea might be to stop the diffusion process completely for gradients beyond the contrast sensitivity. This can be accomplished by *cutting* off the PERONA-MALIK diffusivity, which automatically yields well-posedness results by standard assertions about maximal monotone operators; similar ideas were also applied in [91]. However, all these possibilities are not qualified to enhance edge-like structures in images. This was first achieved by CATTÉ, LIONS, MOREL and COLL in the fundamental paper [39]. The basic idea is simply to *pre-smooth* the diffusivity $g(\cdot)$ by a GAUSSIAN or any other smoothing kernel in order to obtain solutions in the *viscosity* sense. This results in the modified PERONA-MALIK diffusion

$$\mathbf{D}_t \mathbf{u}_t(\vec{x}) = \operatorname{div}_{\vec{x}}(g(|K(\vec{x}) * \nabla_{\vec{x}} \mathbf{u}_t(\vec{x})|) \cdot \nabla_{\vec{x}} \mathbf{u}_t(\vec{x})). \quad (8.8)$$

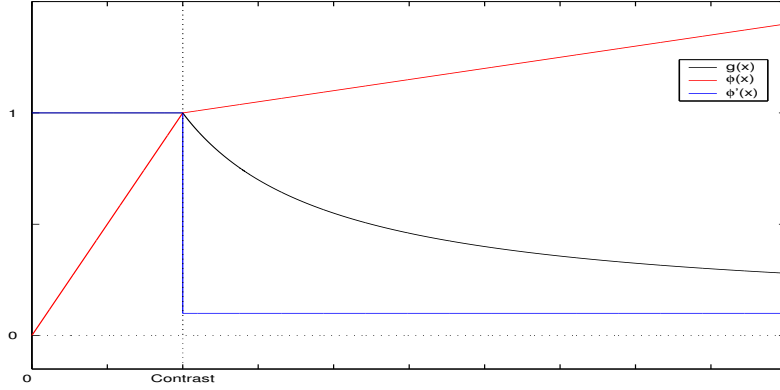


Figure 8.3. SCHNÖRRs diffusivity function $g(\cdot)$, flux $\phi(\cdot)$ and first flux derivative $\phi'(\cdot)$.

As long as the smoothing function $K(\vec{x})$ yields a minimum degree of regularity, existence and uniqueness of a solution for the problem (8.8) were derived in [39]. The enhancement capabilities of this pre-smoothed nonlinear diffusion yield visually impressive regularization results, especially under usage of optimized wavelet filters, see Figure 8.4.

8.2.2. Anisotropic Diffusion

Nonlinear diffusion as presented in the previous paragraph yields an inhomogeneous diffusion over the image range. But the local diffusion in each image point is only depending on the magnitude of the edge gradient and not on the *direction* of the largest variation (or any other significant quantity like e.g. the local coherence). Such a process can be realized by *anisotropic* diffusion. By viewing the considered image as an embedded manifold⁴ to obtain a matrix-valued diffusivity function taking the local orientation of the image into account, as desired; such a proceeding has been proposed by several authors, see e.g. [80] [234] [123] [53].

The mathematical description of anisotropic diffusion looks very similar to that of inhomogeneous diffusion, it is formulated as

$$\mathbf{D}_t \mathbf{U}_t(\vec{x}) = \operatorname{div}_{\vec{x}}(\mathbf{G}(|\nabla_{\vec{x}} \mathbf{U}_t(\vec{x})|) \cdot \nabla_{\vec{x}} \mathbf{U}_t(\vec{x})). \quad (8.9)$$

⁴One widely used metric applied in this situation is the well-known *structure tensor*

$$\mathbf{J} = \nabla_{\vec{x}} \mathbf{U}_t \cdot \nabla_{\vec{x}} \mathbf{U}_t^T.$$

Other choices which are more adapted to color image processing and/or higherdimensional feature spaces are investigated in the work of SOCHEN and KIMMEL [121] [122].



Figure 8.4. Comparison of linear and nonlinear diffusion methods. *Upper left.* Original image. *Upper right.* Linear scale space at scale $t = 2$. *Lower left.* Regularized PERONA-MALIK diffusion using finite differences. *Lower right.* Regularized PERONA-MALIK diffusion using optimized nonseparable wavelets. Parameters for the two examples in the lower row are the same, namely $\lambda = 1$ and $t = 10000$. Implementation characteristics are discussed in Section 8.3.

The main difference to the isotropic nonlinear situation is that a matrix-valued diffusivity $\mathbf{G}(|\nabla_{\vec{x}} \mathcal{U}_t(\vec{x})|)$ stemming from the induced RIEMANNian metric is utilized. To give a physical interpretation, this gives a heat diffusion over the surface of the considered manifold; this allows to *steer* the image diffusion process with respect to local image structures as it was realized by WEICKERT yielding *edge-enhancing* and *coherence-enhancing* anisotropic diffusion [236]. By adapting the techniques of [39], WEICKERT could show that anisotropic diffusion processes like these can be formulated in a well-posed manner if a pre-smoothing as in (8.8) is applied to (8.9).

While anisotropic diffusion is a theoretical step ahead, its practical value depends on the application one has in mind. Regarding e.g. the problem of edge-enhancing (or edge-recovering), anisotropic diffusion leads to a better noise reduction along edges due to its directed diffusion, but on the other hand, more complicated image structures like corners also get diffused (and thus rounded) by this process. Such structures are better recovered by inhomogeneous nonlinear diffusion since only the absolute value of the image gradient is taken into account. But the improved structure recovering capabilities of undirected diffusion have to be paid by lower noise reduction performance. Some results that demonstrate these facts can be seen for example in [236], p. 117, Figure 5.2 and p. 121, Figure 5.4.

8.2.3. Nonlinear Wavelet Estimation

The PDE based methods for nonlinear scale space embeddings presented in the previous sections are one possibility for realizing *information reduction* in images. In the following, we will describe a more functional analytic alternative depending on the embedding of given images into BESOV spaces. BESOV spaces are especially well-suited to this approach since they have an information-theoretic interpretation that fits well to the aspect of reduction of information [14] and are moreover known to exhibit a simple description in the wavelet framework. This technique has been successfully applied in image denoising and optimal lossy compression of image data [77] [78] [41]. Given an image $I(\vec{x})$, $\vec{x} \in \Omega$, the problem is to find a new image $U(\vec{x})$ such that the functional

$$F(U) = \|I(\vec{x}) - U(\vec{x})\|_{L_2(\Omega)} + \lambda \cdot \|U(\vec{x})\|_{S(\Omega)}, \quad \lambda > 0 \quad (8.10)$$

attains a global minimum for all possible $U(\vec{x})$ contained in the considered function space $S(\Omega) \cap L_2(\Omega)$. While the first term preserves the *closeness* to the original image, the second expression forces the result to belong to the (in some sense) more regular function space $S(\Omega)$. The factor λ fixes the weight between both requirements. Suppose now that $S(\Omega)$ is taken to be some BESOV space, that is

$$S(\Omega) = B_{p,q}^s(\Omega).$$

Choosing wavelet expansions as in (5.13) for $I(\vec{x})$ (expansion coefficients are denoted by $c_k^{j,r}$) and $U(\vec{x})$ (coefficients $d_k^{j,r}$), the problem of minimizing (8.10) comes essentially down to

$$\min! \left(\sum_{j,\vec{k},r} |c_k^{j,r} - d_k^{j,r}|^2 + \lambda \cdot \left(\sum_j \left(\sum_{\vec{k},r} |\det \mathbf{Q}|^{j \cdot (p-2) + j \cdot s \cdot p} \cdot |d_k^{j,r}|^p \right)^{q/p} \right)^{1/q} \right), \quad (8.11)$$

where we used the BESOV space characterization (B.11) for functions given by means of a wavelet expansion. For the choice $p = q$ (i.e. the generalized SOBOLEV spaces), the scales in (8.11) are completely decoupled leading to the simple single problems

$$\min! \left(|c_{\vec{k}}^{j,r} - d_{\vec{k}}^{j,r}|^2 + \lambda \cdot |\det \mathbf{Q}|^{j \cdot (p-2) + j \cdot s \cdot p} \cdot |d_{\vec{k}}^{j,r}|^p \right) \quad (8.12)$$

for each \vec{k}, j, r . This latter expression can always be minimized up to arbitrary numerical precision by simple calculus giving the desired approximation of the original image. Moreover, this minimization is especially simple, since it only requires the direct comparison of the coefficients $c_{\vec{k}}^{j,r}$ and $d_{\vec{k}}^{j,r}$ — no *complicated* calculations have to be carried out. However, even for a coupled scale representation, that is for $p \neq q$ in (8.11), the minimization problem can always be solved approximately by algorithms like the classical LEVENBERG-MARQUARDT approach [151].

The question for the *best* choice of the BESOV space parameters naturally arises in this situation. In [41], it was shown how nonlinear wavelet parameter estimations like (8.11) can be used to achieve optimal denoising results by choosing the smoothness parameter s according to the BESOV regularity of the original image [75] and taking $p = q$ with $1/q = s/2 + 1/2$. On the other hand, OSHER and RUDIN have demonstrated the importance of the space $BV(\Omega)$ of bounded maximum variation in the context of image restoration [197]. This space has the nice property to allow *jumps* to be contained in the image data — an important feature for recovering edges and textures. Unfortunately, the space $BV(\Omega)$ has no simple description in the wavelet domain, which makes simple estimations like (8.11) inaccessible. But by validity of Lemma B.5, we may at least *approximate* $BV(\Omega)$ by means of BESOV spaces:

$$B_{1,1}^1(\Omega) \subset BV(\Omega) \subset B_{1,\infty}^1(\Omega).$$

While $B_{1,1}^1(\Omega)$ is contained in $BV(\Omega)$, it contains only continuous functions, or in other words, estimation with respect to $B_{1,1}^1(\Omega)$ can never recover edges or edge-like structures. Anyway, in [93] it is shown that not even the space $BV(\Omega)$ is unconditionally suited to represent natural images. Consequently, we may still apply BESOV space approximation and reduce the smoothness parameter s in order to be able to recover edge-like structures. In Figure 8.5, some results for wavelet estimation in BESOV spaces are compared.

These examples show that nonlinear wavelet estimations give indeed regularized versions of the original images over a range of smoothness scales. If the smoothness parameter is chosen too small (or even negative) this method is



Figure 8.5. Results of nonlinear wavelet estimation. *Upper left.* Original Barbara test image. *Upper right.* Wavelet estimated version in the space $B_{1,1}^1(\Omega)$ with weighting factor $\lambda = 20$. *Lower left.* Same in the space $B_{1,1}^0(\Omega)$. *Lower right.* Estimation in the space $B_{1,1}^2(\Omega)$ — the decoupled smoothing effect on different scales corresponding to the formulation (8.12) becomes evident.

no longer capable of removing noise from images,⁵ but for well chosen s , also the denoising results are quite satisfying, see Figure 8.6.

We conclude that this *nonlinear wavelet scale space* is a good candidate for a simplifying preprocessing step, since it obeys a (steerable) regularizing property and stays *close* to the original image simultaneously. But since the underlying function space characterization can not process special local geometric structure in contrast to the differential geometrically motivated nonlinear diffusion methods presented in the previous paragraphs, these latter ones are superior when it comes to more vision oriented tasks that require the correct recovery of boundaries or local image structures. Consequently, in the upcoming optical

⁵Seen from a functional analytical point of view, noise is nothing else but a distribution with negative regularity index. Consequently, it cannot be eliminated if the smoothness index s is of the same low order.



Figure 8.6. Results of nonlinear wavelet denoising. *Left.* Original image. *Middle.* Noisy version. *Right.* Result of nonlinear wavelet estimation. *Upper row.* Denoising in $B_{1,1}^1(\Omega)$, $\lambda = 30$, PSNR = 32.24. *Lower row.* Denoising in $B_{1,1}^0(\Omega)$, $\lambda = 30$, PSNR = 29.38. Note the remaining of noise in the case where s is chosen too small.

flow calculations, we will apply nonlinear wavelet estimation as a preprocessing and postprocessing step and use nonlinear diffusion methods to regularize the results during the flow calculations itself. To close this section, we just mention a different wavelet-based technique for nonlinear regularization, namely the multiscale *edge reconstruction* proposed in [149]. The key idea is to keep only those edges which persist over a fixed number of scales in the wavelet domain and to remove all other ones; similarly to the methods presented above, *cartoon-like* versions of the images are obtained as a result of such a process. But this method requires a bigger computational amount than nonlinear estimation in BESOV spaces and is thus not used in our implementations.

8.3. Implementation of Nonlinear Diffusion

It was shown above that nonlinear diffusion processes can be realized in a very efficient manner under application of the additive operator splitting method. Typically, this is implemented using finite difference (or finite element) methods leading to the problem of solving $k \cdot n_1 \cdot n_2$ tridiagonal systems of size

$n_1 \times n_1$ or of size $n_2 \times n_2$ respectively instead of one huge $(n_1 \cdot n_2) \times (n_1 \cdot n_2)$ -system having far off-diagonal entries [237]. The purpose of this section is to explain how the wavelet framework together with the linear wavelet scale space can be employed to decrease the computational amount required for nonlinear diffusion further.

Recalling (8.8), nonlinear diffusion was regularized by a pre-smoothing of the image gradient in order to obtain a sufficiently smooth diffusivity that keeps the process stable. Now, our idea is the following: by a semi-implicit discretization scheme, we may decompose the whole process into the predict part which shall be given by the smoothed gradient diffusivity and the update part given by the original image gradient, cf. (7.15). Since the linear wavelet scale space embedding introduces regularization of the original data, it is well-suited for the required pre-smoothing of the diffusivity. This embedding can also be viewed as a projection of the original diffusivity onto some scale j followed by backprojection to the scale 0. If we could at least approximately commute the order of backprojection and application of the diffusivity function $g(\cdot)$, it would be possible to calculate the inverse operator for the predict part according to (7.19) on the j -th scale instead of on scale 0. Since the data on the j -th scale is by a factor of $|\det \mathbf{Q}|^j$ smaller than the original data, this would lead to inversions of correspondingly smaller equation systems and improve the processing speed by another order of magnitude. The working way of this idea is visualized in Figure 8.7

Now, we will show that there is indeed a suited approximation to do the work on a scale j instead of the scale 0. To demonstrate this, the wavelet coefficients with respect to the space V_0 for a given function f may be denoted by $f_{\vec{k}}^0$. Then, by (4.2) and (5.14), the expansion coefficients on some coarser scale j satisfy the estimation

$$f_{\vec{k}}^0 \approx |\det \mathbf{Q}|^{-j/2} \cdot f_{\mathbf{Q}^{-j} \cdot \vec{k}}^j.$$

Hereby, the approximation quality may be measured in terms of the local regularity of the image data, since the coefficients on the higher scales are obtained via averaging those on the lower scales and the variation of neighbouring wavelet coefficients is indeed a regularity measure for functions [75]. Moreover, a number of vanishing moments also guarantees approximation of polynomial order. Anyway, the estimate above motivates us to use the approximation

$$\mathbf{P}_0 \circ \mathbf{P}_j(g(f(\vec{x}))) \approx |\det \mathbf{Q}|^{j/2} \cdot \mathbf{P}_0(g(|\det \mathbf{Q}|^{-j/2} \cdot \mathbf{P}_j(f)(\vec{x}))) \quad (8.13)$$

instead of the correct smoothed diffusivity in nonlinear diffusion and consequently, the inversion to obtain the predict part is done on scale j .

$$\begin{aligned}
 & \left(\begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} - \tau \cdot \begin{bmatrix} \times & & \\ & \times & \\ & & \times \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} \text{Lena} \end{bmatrix} \\
 & \left(\begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} - \tau \cdot \begin{bmatrix} \times & \times & \times \\ & \times & \times \\ & & \times \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} \text{Lena} \end{bmatrix} \\
 & \left(\begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix} - \tau \cdot \begin{bmatrix} \times & & \\ & & \\ & & \end{bmatrix} \right)^{-1} \cdot \begin{bmatrix} \text{Lena} \end{bmatrix}
 \end{aligned}$$

Figure 8.7. Schematic illustration of wavelet-based additive operator splitting and its reduced complexity. The expression $(\mathbf{Id} - \mathbf{N}(f) \circ \mathbf{L})^{-1} \cdot f$ (see 7.15) is sketched for various formulations. The formulation in the first two rows is equivalent for orthogonal or tight frame wavelet transforms.⁶ The third row describes the regularized approximation within the linear wavelet scale space by means of (8.8) and (8.13). The inversion in the third row is obviously simpler than in the first two.

The reader should be warned that the choice of scale on which one approximates is of crucial significance — a too large chosen scale may totally distort the original image, while a too small chosen scale would keep undesired artefacts like noise. On the other hand, the same may also happen if one applies the *classical* linear scale space using a GAUSSIAN of *unsuited* standard deviation. In [242], the dependence between the scale parameter and the evolution time for the regularized PERONA-MALIK diffusion process have been investigated, leading to the conclusion that the scale parameter should decrease with increas-

⁵This can be seen as follows:

$$\begin{aligned}
 \mathbf{P}_0(\mathbf{A}^{-1} \cdot f) &= \mathbf{U} \cdot \mathbf{A}^{-1} \cdot f \cdot \mathbf{U}^T \\
 &= \mathbf{U} \cdot \mathbf{A}^{-1} \cdot \mathbf{U}^T \cdot \mathbf{U} \cdot f \cdot \mathbf{U}^T \\
 &= \mathbf{U}^{-T} \cdot \mathbf{A}^{-1} \cdot \mathbf{U}^{-1} \cdot \mathbf{U} \cdot f \cdot \mathbf{U}^T \\
 &= (\mathbf{U} \cdot \mathbf{A} \cdot \mathbf{U}^T)^{-1} \cdot \mathbf{U} \cdot f \cdot \mathbf{U}^T \\
 &= (\mathbf{P}_0(\mathbf{A}))^{-1} \cdot \mathbf{P}_0(f)
 \end{aligned}$$

where \mathbf{U} is the orthogonal transformation matrix of the wavelet transform, \mathbf{A} is the matrix representation of the inner operator on the chosen decomposition scale and f denotes the (image) data.

ing evolution time. This decreasing of scale reflects the increasing reliability of the processed image during the temporal evolution. Following the results of [242], we typically calculate the first few evolution steps on the scale $j = 3$ or $j = 4$ (depending on the size of the original image) and decrease j subsequently. So the reduction of computational costs gets smaller with proceeding time, but the method remains still faster than *standard* additive operator splitting by all means.

Chapter Summary

In Chapter 8, we have recalled the classical axiomatics of linear scale spaces leading to a unique solution, namely the convolution with a GAUSSian. Then, we have shown how a slight relaxation of one of the axioms made different linear scale spaces available. In particular, we have introduced a linear wavelet scale space that can be derived from any symmetric, unimodal and nonnegative scaling filter. The similarities and the differences between these two types of scale spaces were demonstrated afterwards. In order to realize image enhancing properties, a number of nonlinear scale spaces were investigated in the second part of this chapter. As in the first section, a wavelet-based generalization of the original nonlinear scale space concept was introduced. Here, this generalization was done by an embedding of the original data into certain function spaces of BESOV type. We have chosen BESOV spaces for this embedding, since they have already shown to be well-performing in the context of denoising and due to their special information-theoretical interpretation. The chapter was closed by the presentation of a new implementation of nonlinear diffusion making use of additive operator splitting and also maintaining the special sparsity structure of the image representation in the new linear wavelet scale space.

Chapter 9

OPTICAL FLOW

Everything is in flow.

—HERAKLIT

The present chapter deals with the computation of the optical flow occurring within a given sequence of images. After a short introduction into the topic of consideration, several basic models for the evaluation of optical flow are presented and discussed. In the following section, the proposed flow model is studied in a more detailed way. In particular, its mathematical well-posedness and its efficient numerical implementation are investigated. Afterwards, some possible further model extensions covering various practical aspects are investigated and their mathematical formulation within the proposed framework is sketched. Section 9.5 mainly deals with several examples. Results for different test scenarios are evaluated and compared to other approaches. The actual chapter is closed by a juxtaposition of the presented methods to other methods found in the literature.

9.1. Introduction

The reliable computation of the optical flow is one of the fundamental problems in the processing of image sequences, since it is supposed to give an approximation to the *real image motion*. Hereby, image motion is meant to be the projection of the velocities of the considered subset of the fourdimensional space-time onto a three-dimensional image-plane-time fixed by the visual sensor and the time interval taken into account. In many practical situations, optical flow turns out to be a quite good approximation of image motion, however, both notions are not equivalent. While optical flow measures the apparent *visual changes* in image sequences, image motion gives the *real geometric*

transformation field caused by motion within the considered scene.¹ However, since we usually have only the projection of the scene, it is generally impossible (even with a multi-camera system) to detect image motion, while the optical flow is accessible from the recorded visual information.

The computation of optical flow is one of the key tasks in computer vision and finds a wide variety of applications. Examples range from object tracking, robot navigation and autonomous driving systems to efficient algorithms for video coding or stereo matching. Since the data to be processed is at least three-dimensional — thus quite huge in general — and most of the abovementioned (as well as other) applications are usually very time-critical, the need for fast *and* reliable algorithms is still the most challenging task for optical flow calculations. The wavelet framework has proven to be very efficient for many other image processing tasks (see e.g. the previous chapter) and consequently, we will develop a suited wavelet-based optical flow framework in this chapter.

9.2. Models for Optical Flow

There are a set of different methods to calculate the optical flow. Before starting with our considerations, we shall briefly categorize these techniques with respect to the underlying mathematical concepts to better understand the technical mechanisms behind.

Maybe the most popular approach to optical flow are *differential techniques*. These mainly rely on the computation of the flow field from spatiotemporal derivatives of the image intensity function (or preprocessed versions of it, see below) and a geometric model setting the image derivatives and the sought flow field into some mathematical relation, as it will be done below in (9.2). It is directly evident that the considered image sequences must be at least differentiable for any differential technique to make sense at all. This is (mostly) achieved by spatiotemporal presmoothing of the original data. In the most differential approaches, additional constraints like regularizing terms and/or local averaging are applied in order to improve the reliability of the estimations. Due to the local character of differential methods, a multiscale strategy is useful to detect *fast* motions. Influencing works in the area of differential optical flow are e.g. [114] [165] [24] [233] [86] and [5], some of these will be discussed in

¹The most striking example to demonstrate the difference between optical flow and image motion is that of a rotating, ideally homogeneous sphere. In this situation, there is no optical flow, since the homogeneity of the sphere does not cause any changes in the visual patterns, but there is indeed motion, namely the assumed rotation. Contrarily, we may also consider the same sphere now standing still and start to point a light source onto the sphere. If this light source is moved, there occurs optical flow, since the visual pattern changes. But this time, we don't have any real image motion, because the sphere was assumed to stand still. However, artificial situations like the depicted ones appear rather exceptional in practice.

Section 9.6 in the context of comparing the results achieved in this work with those of other researchers.

The second large class of approaches is that of matching algorithms. Hereby, a *similarity measure* as e.g. the minimum distance or cross-correlation is applied to temporal neighbouring image frames in order to find the best fit between the considered frames. Matching techniques do not suffer from differentiability issues and are capable to handle fast motions since they are not depending on the local structure of the images. On the other hand, they are computationally much more expensive, because the non-locality leads to large (full) equation systems. Fortunately, application of coarse-to-fine strategies can help to reduce the computational cost considerably. Matching techniques are not considered in this work, the interested reader will find more material on the topic in [7] [209] or [247] and the references therein.

Other techniques to recover the optical flow are *energy-based* or *phase-based* methods. These stem from reformulations of the original problem in the FOURIER domain involving very different features. Since these approaches also require very special techniques, a classification would be too lengthy to be done here. More about such approaches can be found in the survey article [15] and the literature cited there.

9.2.1. A Differential Flow Model

To start with the derivation of suited models to compute the optical flow, it is sensible to begin with the simplest possible framework. As an initial assumption, it is very common to restrict the model by the *brightness constancy constraint*. This means to assume that the measured intensity of an image point remains invariant over time along his motion trajectory in the image sequence domain. It is formulated by means of

$$I(\vec{x}(t), t) = \text{const.} \quad (9.1)$$

where $I(\vec{x}, \cdot) \in L_2(\Omega \times T)$ shall denote the image intensity function and the sets $\Omega \subset \mathbb{R}^2$, $T \subset \mathbb{R}$ are the image plane and the considered time interval respectively. Expanding expression (9.1) into a simple TAYLOR series directly yields

$$\langle \nabla_{\vec{x}} I(\vec{x}(t), t), \vec{w}(\vec{x}, t) \rangle + \mathbf{D}_t(I)(\vec{x}(t), t) \approx 0 \quad (9.2)$$

where the higher order expansion terms are neglected for simplicity and

$$\vec{w}(\vec{x}, t) = [u(\vec{x}, t), v(\vec{x}, t)]^T = [\mathbf{D}_t(x_1)(t), \mathbf{D}_t(x_2)(t)]^T$$

is the sought optical flow vector field. Equation (9.2) is the starting point for many differential approaches to the problem. Unfortunately, there are several

problems coming along with this optical flow model, which turns out to describe an ill-posed problem. First, the constraint equation (9.2) is obviously underdetermined and therefore, it is not possible to compute the whole flow information out of this equation. To circumvent this, the simplest approach is to apply a *local weighting* of the information in order to get more constraints. This usually leads to an overdetermined system that has to be optimized by least squares or total least squares methods [233]. However, if the local structure of the considered image frames is intrinsically onedimensional, even a local weighting of the visual information remains insufficient to solve (9.2). In this situation, only the normal part of the flow field is extractable² — this phenomenon is well-known as the *aperture problem*, its occurrence is sketched in Figure 9.1. The aperture problem cannot be completely overcome, since the motion of a homogeneous contour is ambiguous [166] [167], but the addition of *non-local* constraints is usually very helpful since homogeneous contours in real-world images do generally not cover the whole image plane but only local parts of it.

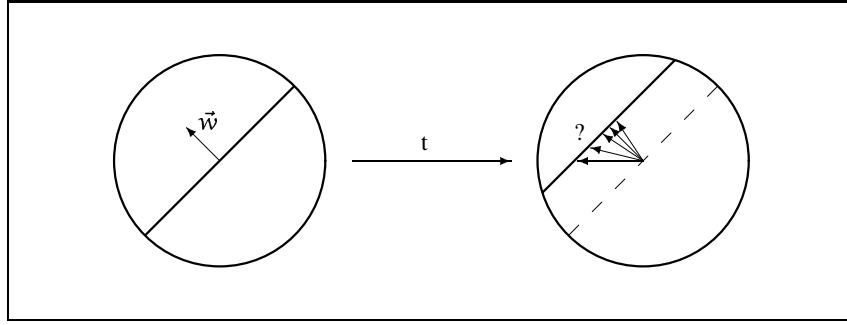


Figure 9.1. The aperture problem — only the normal part of the velocity vector \vec{w} is uniquely determined.

There are more inadequacies with the model (9.2) than the aperture problem. The local averaging of the patterns of visual information helped to solve the aperture problem at least for small local regions, but this proceeding comes along with a blurring effect at motion boundaries caused by the spatial overlapping of information. This undesired effect gets even more intensified by presmoothing the image frames in order to guarantee their differentiability and

²Obviously, the normal share \vec{w}_n of the flow field targets at the same direction as the image gradient $\nabla_{\vec{x}} I(\vec{x}(t), t)$ and consequently, \vec{w}_n may be written as

$$\vec{w}_n = -\|\nabla_{\vec{x}} I(\vec{x}(t), t)\|_{L_2}^{-2} \cdot \mathbf{D}_t(I)(\vec{x}(t), t) \cdot \nabla_{\vec{x}} I(\vec{x}(t), t)$$

by (9.2) and a simple geometric argument.

make them just fit into the chosen framework. Furthermore, by assuming (9.1), we implicitly took the reflectance of the considered scene as well as its illumination to be constant over the whole time interval $T \subset \mathbb{R}$ taken into account. These are very restrictive assumptions which are by no means maintainable for typical real world situations and should be weakened by suited modifications of the optical flow model. This is the point of consideration in the upcoming paragraphs of this section.

9.2.2. Illumination Changes

In order to form our flow model in such a way that it is capable to deal with *realistic* lightning conditions (moving shadows, appearing and disappearing of light sources, dimming or diffusion of light by clouds, ...), the brightness constancy assumption has to be exchanged by a somewhat more suited a priori specification. Consequently, one should consider not only geometric information as in (9.1), but also radiometric information as sources to compute the optical flow. The proximate way to model such dynamic illumination is to decompose the image sequence into the factors of a *reflectance* part R and a *luminance* part λ as well as an additional *offset* c yielding

$$I(\vec{x}(t), t) = R(\vec{x}(t), t) \cdot \lambda(\vec{x}(t), t) + c(\vec{x}(t), t).$$

Assuming that the reflectance of an image point does not change significantly along his motion path, $R(\vec{x}(t), t)$ must satisfy (9.2). Furthermore, it is natural to postulate that the luminance and the offset vary slowly in space, since light is in general distributed locally homogeneous. This requirement is formulated mathematically by $\nabla_{\vec{x}}\lambda = \vec{0}$ and $\nabla_{\vec{x}}c = \vec{0}$, yielding the modified optical flow equation

$$F(\vec{w}, l, c) = \langle \nabla_{\vec{x}}I, \vec{w} \rangle + \mathbf{D}_t(I) - I \cdot l - c \approx 0, \quad (9.3)$$

where l abbreviates the logarithmic time derivative $l = (\mathbf{D}_t\lambda)/\lambda$ keeping the notation as simple as possible. This type of reformulation was first proposed in [174]. From a mathematical point of view, it would be sufficient to use either luminance or offset in the model instead of both entities [173], but from a radiometric point of view, the combined usage of both parameters is more natural, since they model different aspects of illumination. Multiplicative luminance is usually found in source illumination or at reflecting surfaces, while additive light occurs in the case of moving shadows or spots. Therefore, usage of both parameters often leads to better flow estimations and additionally, it turned out to be numerically stabilizing in our experiments. Anyway, in [19], an algorithm relying solely on luminance changes was presented and noticeably improved optical flow results for image sequences with brightness changes were reported there.

Although the model (9.3) gives a complete mathematical description of all possible types of illumination changes, it might be generalized further. Such generalizations are usually targeting at special applications that do not only require the flow estimations itself but also the calculation of additional parameters in order to extract certain *processes* from the observed scene. An example of such a process is physical oceanography as considered in [108]. There, physical brightness change models are applied in order to evaluate three-dimensional hydrodynamic undersea transport processes. Anyway, since we are *only* looking for fast and reliable computation methods for the optical flow and not for the additional recovery of physical processes, we do not require such generalized illumination models and limit ourselves to the mathematically sufficient relation (9.3).

9.2.3. Nonlinear Regularization

Spatiotemporal presmoothing of the image data and local averaging of the visual information are important ingredients to obtain sensible solutions for the optical flow problem. However, as was argued before, these cause a spatial overlapping of the patterns of information which leads to blurring effects. Such effects are harmless in regions of relative homogeneity, but in textured areas and especially at motion boundaries, the smoothing effect is undesired and may destroy at least parts of important flow information. Namely, for many applications, it is just the exact recovery of these motion boundaries which is of high importance.

Recalling the regularization results presented in the previous chapter, the framework of nonlinear diffusion offers a way to regularize the optical flow problem in such a way, that (motion) boundaries may be recovered or even enhanced. This edge-enhancement property compensates the presmoothing of the data which is necessary to make the differential formulation sensible. In analogy to the relations (8.2) and (8.5), nonlinear diffusion of the optical flow field may be achieved by the evolution process

$$\begin{aligned} \mathbf{D}_\tau \vec{u}(\vec{x}, t)(\tau) &= \operatorname{div}_{\vec{x}} \left(g(|\nabla_{\vec{x}} \vec{u}(\vec{x}, t)(\tau)|^2 + |\nabla_{\vec{x}} \vec{v}(\vec{x}, t)(\tau)|^2) \cdot \nabla_{\vec{x}} \vec{u}(\vec{x}, t)(\tau) \right) \\ \mathbf{D}_\tau \vec{v}(\vec{x}, t)(\tau) &= \operatorname{div}_{\vec{x}} \left(g(|\nabla_{\vec{x}} \vec{u}(\vec{x}, t)(\tau)|^2 + |\nabla_{\vec{x}} \vec{v}(\vec{x}, t)(\tau)|^2) \cdot \nabla_{\vec{x}} \vec{v}(\vec{x}, t)(\tau) \right). \end{aligned} \quad (9.4)$$

From a physical point of view, it might be a good idea to apply similar nonlinear diffusion processes also to the luminance and the illumination offset as well yielding

$$\begin{aligned} \mathbf{D}_\tau \vec{l}(\vec{x}, t)(\tau) &= \operatorname{div}_{\vec{x}} \left(g(|\nabla_{\vec{x}} \vec{l}(\vec{x}, t)(\tau)|^2) \cdot \nabla_{\vec{x}} \vec{l}(\vec{x}, t)(\tau) \right) \\ \mathbf{D}_\tau \vec{c}(\vec{x}, t)(\tau) &= \operatorname{div}_{\vec{x}} \left(g(|\nabla_{\vec{x}} \vec{c}(\vec{x}, t)(\tau)|^2) \cdot \nabla_{\vec{x}} \vec{c}(\vec{x}, t)(\tau) \right). \end{aligned} \quad (9.5)$$

In addition to these evolutions, one has to assure that the optical flow constraint (9.3) still holds or reaches at least a minimum value in order to receive credible flow estimations. This is realized by the simultaneously minimization of the *optical flow energy functional*

$$\int_{\Omega} F(u, v, l, c)^2 d\vec{x} = \min!$$

where $F(u, v, l, c)$ is as defined in (9.3). From calculus we know that a necessary condition for this functional to attain a minimum value is the fulfillment of the relations (vanishing of the first variations)

$$\mathbf{D}_u F = 0 \quad \mathbf{D}_v F = 0 \quad \mathbf{D}_l F = 0 \quad \text{and} \quad \mathbf{D}_c F = 0.$$

Linking these requirements with the evolution processes (9.4) and (9.5) leads to the coupled diffusion-reaction system

$$\begin{aligned} u_{\tau} &= \alpha \cdot \operatorname{div} \left(g(|\nabla_{\vec{x}} u|^2 + |\nabla_{\vec{x}} v|^2) \cdot \nabla_{\vec{x}} u \right) \\ &\quad - I_{x_1} \cdot (I_{x_1} \cdot u + I_{x_2} \cdot v + I_t - I \cdot l - c) \\ v_{\tau} &= \alpha \cdot \operatorname{div} \left(g(|\nabla_{\vec{x}} u|^2 + |\nabla_{\vec{x}} v|^2) \cdot \nabla_{\vec{x}} v \right) \\ &\quad - I_{x_2} \cdot (I_{x_1} \cdot u + I_{x_2} \cdot v + I_t - I \cdot l - c) \\ l_{\tau} &= \beta \cdot \operatorname{div} \left(g(|\nabla_{\vec{x}} l|^2) \cdot \nabla_{\vec{x}} l \right) \\ &\quad + I \cdot (I_{x_1} \cdot u + I_{x_2} \cdot v + I_t - I \cdot l - c) \\ c_{\tau} &= \gamma \cdot \operatorname{div} \left(g(|\nabla_{\vec{x}} c|^2) \cdot \nabla_{\vec{x}} c \right) \\ &\quad + I_{x_1} \cdot u + I_{x_2} \cdot v + I_t - I \cdot l - c, \end{aligned} \tag{9.6}$$

where α , β and γ are free weighting parameters that allow to steer whether the process is more flow recovering or more flow (and illumination) regularizing. These parameters may also be interpreted as LAGRANGE multipliers of the equivalent variational formulation of this problem. Anyway, the *best* choice of this parameters is an experimental matter and depends strongly on the type of image data.

Relations (9.6) are the basic optical flow model we propose for fast and reliable calculations. Due to the incorporation of multiplicative and additive illumination changes and nonlinear diffusion processes, it gives very good results for some of the standard test sequences (see also Section 9.5). Nevertheless, this model is by no means *complete* in the sense of feasibility of any possible visual occurrence. Consequently, some more involved extensions to (9.6) will be discussed in an upcoming section. At this stage, we only point out that there would be two *obvious* possibilities to extend (9.6) in a straightforward way we won't make use of. For the first, this would be the incorporation of anisotropic

diffusion as defined in Section 8.8.2. We will not use anisotropic diffusion due to its poorer capabilities to recover corner-like structures in comparison to inhomogeneous diffusion. The improved noise-reduction of anisotropic diffusion is of minor interest for us since we will apply a nonlinear wavelet-based preprocessing in our implementations and thus assume our data to be (more or less) denoised. The second extension one might directly think of is a full spatiotemporal processing of the model (9.6) as it was proposed in [238]. Spatiotemporal regularization has proven to smoothen out noise and to preserve motion boundaries better than pure spatial regularization. However, such a proceeding is not *realistic* for typical applications, because the situation that the whole movie is available at once is a rather exceptional case in practice. Usually, one has only a few frames disposable at each time step and in this situation, spatiotemporal processing is not feasible. Therefore, we renounce on this extension possibility in our investigations.

9.3. Analytical and Numerical Investigation of the Problem

This section is devoted to the deeper mathematical study of the model (9.6) from the well-posedness and from the implementational point of view. Existence and uniqueness results for the optical flow problem under certain regularity prerequisites are proven in the first paragraph. A thorough discussion of the efficient wavelet-based implementation of the process (9.6) will be given in the second paragraph.

9.3.1. Consistency of the Model

This paragraph is dedicated to demonstrate the consistency of the proposed optical flow model. In particular, we will show existence and uniqueness of the solution under some regularity assumptions. For the sake of simple notations, we will denote partial derivatives by the short form

$$\mathbf{D}_{\vec{x}_i}(f)(\vec{x}) = f_{\vec{x}_i}$$

and abbreviate the multidimensional SOBOLEV spaces by

$$\mathbf{W}_p^{q,(m)}(G) = \underbrace{\mathbf{W}_p^q(G) \times \mathbf{W}_p^q(G) \times \cdots \times \mathbf{W}_p^q(G)}_{m \text{ times}}$$

with $G \subset \mathbb{R}^n$. Furthermore, let

$$\mathcal{C}([a, b]; \mathbf{W}_p^{q,(m)}(G)), \quad [a, b] \subset \mathbb{R}$$

be the space of continuous functions in the interval $[a, b]$ having their range in $\mathbf{W}_p^{q,(m)}(G)$. With this prerequisites, we can state and prove the main result of this section.

THEOREM* 9.1 *Let $I(\vec{x}, t) \in W_2^1(\Omega \times T)$ with $\Omega \subset \mathbb{R}^2$ and $T \subset \mathbb{R}$ be a given image sequence and denote by*

$$\vec{s}_0(\vec{x}, t) = (u_0, v_0, l_0, c_0)^T(\vec{x}, t) \in W_2^{2,(4)}(\Omega \times T)$$

any sufficiently smooth initial state. Furthermore, the partial derivatives of the image sequence with respect to space and time are labelled by I_{x_i} and I_t respectively. Then, the optical flow model given by the system (9.6) possesses a unique solution

$$\vec{s}(\vec{x}, t)(\tau) = (u, v, l, c)^T(\vec{x}, t)(\tau) \in \mathcal{C}([0, \infty); W_2^{2,(4)}(\Omega \times T)).$$

As before, (u, v) denotes the optical flow field, l is the measured luminance and c is the illumination offset of the image sequence.

Remark. One might also prove the existence of weak solutions for (9.6) under relaxed regularity constraints on $I(\vec{x}, t)$, $\vec{s}_0(\vec{x})$ if a pre-smoothing of the diffusivity similar to the nonlinear diffusion approach in [39] is applied. The interested reader is referred to this article.

Proof. First, we decompose the system (9.6) into a linear part $\mathbf{L}(\vec{s})$ and a nonlinear part $\mathbf{N}(\vec{s})$ by means of

$$\mathbf{L}(\vec{s}) = - \begin{bmatrix} I_{x_1}^2 & I_{x_1} \cdot I_{x_2} & -I_{x_1} \cdot I & -I_{x_1} \\ I_{x_1} \cdot I_{x_2} & I_{x_2}^2 & -I_{x_2} \cdot I & -I_{x_2} \\ -I_{x_1} \cdot I & -I_{x_2} \cdot I & I^2 & I \\ -I_{x_1} & -I_{x_2} & I & 1 \end{bmatrix} \cdot \begin{bmatrix} u \\ v \\ l \\ c \end{bmatrix} \quad (9.7)$$

and

$$\mathbf{N}(\vec{s}) = \begin{bmatrix} \alpha \cdot \operatorname{div}(g(|\nabla_{\vec{x}} u|^2 + |\nabla_{\vec{x}} v|^2) \cdot \nabla_{\vec{x}} u) - I_{x_1} \cdot I_t \\ \alpha \cdot \operatorname{div}(g(|\nabla_{\vec{x}} u|^2 + |\nabla_{\vec{x}} v|^2) \cdot \nabla_{\vec{x}} v) - I_{x_2} \cdot I_t \\ \beta \cdot \operatorname{div}(g(|\nabla_{\vec{x}} l|^2) \cdot \nabla_{\vec{x}} l) + I \cdot I_t \\ \gamma \cdot \operatorname{div}(g(|\nabla_{\vec{x}} c|^2) \cdot \nabla_{\vec{x}} c) + I_t \end{bmatrix}. \quad (9.8)$$

Following the DUHAMEL principle [248], any classical solution to the system (9.6) is given by

$$(\vec{s})(\tau) = e^{\tau \cdot \mathbf{L}} \vec{s}_0 + \int_0^\tau e^{(\tau-\sigma) \cdot \mathbf{L}} \mathbf{N}(\vec{s})(\sigma) d\sigma \quad (9.9)$$

where $e^{\tau \cdot \mathbf{L}}$ denotes the semigroup associated to the infinitesimal generator \mathbf{L} . Since $I \in W_2^1$, \mathbf{L} is obviously a negative semi-definite, bounded linear operator

and the semigroup is analytical in the domain of \mathbf{L} , which yields the *quasi-contraction* norm estimate

$$\|e^{\tau \cdot \mathbf{L}}(\vec{s})\| \leq e^{\omega \cdot \tau} \cdot \|(\vec{s})\|$$

by a classical result from operator theory, namely the HILLE-YOSIDA theorem [179]. Hereby, the exponent ω corresponds to the eigenvalue of \mathbf{L} with the largest real part. For \mathbf{L} , this is given by $\omega = 0$ leading to

$$\|e^{\tau \cdot \mathbf{L}}(\vec{s})\| \leq \|(\vec{s})\|. \quad (9.10)$$

Additionally, we require a LIPSCHITZ-type estimate for the nonlinear operator part $\mathbf{N}(\vec{s})$. Taking $\vec{h} \in W_2^{2,(4)}(\Omega \times T)$ to be given by

$$\vec{h}(\vec{x}, t) = (h^{(u)}, h^{(v)}, h^{(l)}, h^{(c)})(\vec{x}, t),$$

considering only the first component of $\mathbf{N}(\vec{s})$ and inserting the PERONA-MALIK diffusivity function (8.5), we obtain

$$\begin{aligned} & |\mathbf{N}_1(\vec{s}) - \mathbf{N}_1(\vec{s} + \vec{h})| \\ &= \alpha \cdot \left| \operatorname{div} \left(\frac{1}{1 + (|\nabla_{\vec{x}} u|^2 + |\nabla_{\vec{x}} v|^2)/\lambda^2} \cdot \nabla_{\vec{x}} u \right) - \right. \\ & \quad \left. \operatorname{div} \left(\frac{1}{1 + (|\nabla_{\vec{x}}(u + h^{(u)})|^2 + |\nabla_{\vec{x}}(v + h^{(v)})|^2)/\lambda^2} \cdot \nabla_{\vec{x}}(u + h^{(u)}) \right) \right| \\ &= \alpha \cdot \left| \operatorname{div} \left(\theta(\nabla_{\vec{x}} u, \nabla_{\vec{x}} v, \nabla_{\vec{x}} h^{(u)}, \nabla_{\vec{x}} h^{(v)}) \cdot \nabla_{\vec{x}}(u + h^{(u)}) \right) - \right. \\ & \quad \left. \operatorname{div} \left(\frac{1}{1 + (|\nabla_{\vec{x}} u|^2 + |\nabla_{\vec{x}} v|^2)/\lambda^2} \cdot \nabla_{\vec{x}} h^{(u)} \right) \right| \end{aligned}$$

where $\theta(\nabla_{\vec{x}} u, \nabla_{\vec{x}} v, \nabla_{\vec{x}} h^{(u)}, \nabla_{\vec{x}} h^{(v)})$ is the abbreviated form of the difference function

$$\begin{aligned} & \theta(\nabla_{\vec{x}} u, \nabla_{\vec{x}} v, \nabla_{\vec{x}} h^{(u)}, \nabla_{\vec{x}} h^{(v)}) \\ &= g(|\nabla_{\vec{x}} u|^2 + |\nabla_{\vec{x}} v|^2) - g(|\nabla_{\vec{x}}(u + h^{(u)})|^2 + |\nabla_{\vec{x}}(v + h^{(v)})|^2) \\ &= \frac{\lambda^2 \cdot (|\nabla_{\vec{x}}(u + h^{(u)})|^2 + |\nabla_{\vec{x}}(v + h^{(v)})|^2 - |\nabla_{\vec{x}} u|^2 - |\nabla_{\vec{x}} v|^2)}{(\lambda^2 + |\nabla_{\vec{x}} u|^2 + |\nabla_{\vec{x}} v|^2) \cdot (\lambda^2 + |\nabla_{\vec{x}}(u + h^{(u)})|^2 + |\nabla_{\vec{x}}(v + h^{(v)})|^2)}. \end{aligned}$$

By simple calculus and some case distinctions, it is now possible to verify the estimates

$$\begin{aligned} |\theta(\nabla_{\vec{x}} u, \nabla_{\vec{x}} v, \nabla_{\vec{x}} h^{(u)}, \nabla_{\vec{x}} h^{(v)})| &\leq 1 \quad \text{and} \\ |\theta(\nabla_{\vec{x}} u, \nabla_{\vec{x}} v, \nabla_{\vec{x}} h^{(u)}, \nabla_{\vec{x}} h^{(v)})| &\leq \frac{|\nabla_{\vec{x}} h^{(u)}|}{|\nabla_{\vec{x}} u|} \end{aligned}$$

for $\lambda > 0$, which yields

$$\left| \mathbf{N}_1(\vec{s}) - \mathbf{N}_1(\vec{s} + \vec{h}) \right| \leq \alpha \cdot \left| \operatorname{div}(\tilde{\mathbf{C}} \cdot \nabla_{\vec{x}} \mathbf{h}^{(u)}) \right|.$$

Regarding the fact that we have $\vec{s} \in \mathbf{W}_2^2(\Omega \times \mathbf{T})$ and that similar estimations also hold for the second, third and fourth component of $\mathbf{N}(\vec{s})$, we finally end up with

$$\left\| \mathbf{N}(\vec{s}) - \mathbf{N}(\vec{s} + \vec{h}) \right\| \leq C \cdot \|\vec{h}\| \quad (9.11)$$

for some suited constant C . This is exactly the announced LIPSCHITZ-type estimate for $\mathbf{N}(\vec{s})$.

In the next step, we will show that the solution representing function according to (9.9) which is given by

$$\psi(\vec{s})(\tau) = e^{\tau \cdot \mathbf{L}} \vec{s}_0 + \int_0^\tau e^{(\tau-\sigma) \cdot \mathbf{L}} \mathbf{N}(\vec{s})(\sigma) d\sigma \quad (9.12)$$

has a fixed point and thus a solution by BANACHs fixed point theorem. In order to establish existence, we define the norm

$$\|(\vec{s})\|_{S(\kappa)} = \sup_{\tau \geq 0} \|(\vec{s})\| \cdot e^{-\kappa \cdot \tau}$$

which induces a new BANACH space $S(\kappa)$. Considering the norm of differences of ψ in $S(\kappa)$ yields

$$\begin{aligned} \left\| \psi(\vec{s}) - \psi(\vec{s} + \vec{h}) \right\|_{S(\kappa)} &\stackrel{(9.12)}{=} \sup_{\tau \geq 0} \left\| \int_0^\tau e^{(\tau-\sigma) \cdot \mathbf{L}} (\mathbf{N}(\vec{s}) - \mathbf{N}(\vec{s} + \vec{h})) d\sigma \right\| \cdot e^{-\kappa \cdot \tau} \\ &\leq \sup_{\tau \geq 0} \int_0^\tau \left\| e^{(\tau-\sigma) \cdot \mathbf{L}} (\mathbf{N}(\vec{s}) - \mathbf{N}(\vec{s} + \vec{h})) \right\| d\sigma \cdot e^{-\kappa \cdot \tau} \\ &\stackrel{(9.10)}{\leq} \sup_{\tau \geq 0} \int_0^\tau \left\| \mathbf{N}(\vec{s}) - \mathbf{N}(\vec{s} + \vec{h}) \right\| d\sigma \cdot e^{-\kappa \cdot \tau} \\ &\stackrel{(9.11)}{\leq} C \cdot \sup_{\tau \geq 0} \int_0^\tau \|\vec{h}\| d\sigma \cdot e^{-\kappa \cdot \tau} \\ &\leq C \cdot \sup_{\tau \geq 0} \|\vec{h}\|_{S(\kappa)} \cdot \int_0^\tau e^{-\kappa \cdot \sigma} d\sigma \cdot e^{-\kappa \cdot \tau} \\ &= \frac{C}{\kappa} \cdot \sup_{\tau \geq 0} \|\vec{h}\|_{S(\kappa)} \cdot (1 - e^{-\kappa \cdot \tau}) \\ &\leq \frac{C}{\kappa} \cdot \sup_{\tau \geq 0} \|\vec{h}\|_{S(\kappa)}. \end{aligned}$$

Choosing $\kappa > C$, the functional $\psi(\vec{s})$ becomes a proper contraction on the space $S(\kappa)$. The existence of a fixed point (\vec{s}) of $\psi(\vec{s})$ follows directly from BANACH's fixed point theorem. Furthermore, the uniqueness of this solution for given *evolutionary boundary conditions*

$$(\vec{s})(\vec{x}, t)(0) \equiv (\vec{s}_0)(\vec{x}, t)$$

is an immediate consequence of GRONWALL's inequality [183] which concludes the proof. \square .

The restriction for the flow field and the illumination parameters to belong to the spaces $W_2^2(\Omega \times T)$ might be perceived as a restrictive condition. But the reader should keep in mind that the solutions to the regularized PERONA-MALIK diffusion are also of the same smoothness order as the applied smoothing kernels $K(\vec{x})$ in (8.8) are. The smoothness restriction on the optical flow in Theorem 9.1 is exactly of the same quality. Recalling the very satisfying results of the regularized PERONA-MALIK process, these restrictions should not hinder successful flow estimations under usage of the model (9.6).

Several well-posedness results for the computation of the optical flow can be found in the literature. In particular, the result of Theorem 9.1 looks rather similar to the existence statements made in [4]. But while the authors of [4] investigated a regularization constraint depending on the local image gradient and being linear in the flow variables, we chose a flow-dependent nonlinear regularization term. This makes some estimations more sensitive and finds expression in the fact that more rigorous regularity requirements are necessary. Other well-posedness results on a model similar to (9.6) (but not taking illumination issues in account) are obtained in [238] [202] under usage of a convex diffusivity function $g(\cdot)$, which directly guarantees existence and uniqueness of global minimizers. A more technical investigation yielding well-posedness results in the space $BV(\Omega)$ of bounded variation was carried out in [9] [10]. While the space of bounded variation is a good candidate for realistic flow fields, the framework given there is also restricted to convex regularizers and therefore, no enhancement of motion boundaries can be expected in that case.

9.3.2. Implementation and Numerics

We shall study the characteristics of the model (9.6) under the aspect of its implementation and numerical behaviour now. At the beginning we are only given a *raw* sequence of images $I(\vec{x}, t) \subset L_2(\Omega \times T)$. Having this, the following steps are carried out.

- 1 **Pre-processing.** First, we have to make sure that we indeed meet the requirements of Theorem 9.1. To guarantee this, a nonlinear wavelet estimation

as described in Paragraph 8.8.2.3 is applied to the image sequence $I(\vec{x}, t)$. Hereby, we choose the BESOV space $B_{1,1}^1(\Omega \times T)$ as embedding space. Since the original data was also in $L_2(\Omega \times T)$, we have in fact

$$I(\vec{x}, t) \subset W_2^1(\Omega \times T)$$

for the pre-processed image sequence, as it is required in Theorem 9.1.

- 2 **Initial guess.** Before the evolution iteration can be started, a *sufficiently good* initial state $\vec{s}_0 = (u_0, v_0, l_0, c_0)^T$ has to be calculated. In order to obtain it, we use the non-regularized model equation (9.3) and apply a local averaging with least squares optimization to circumvent the aperture problem. Hereby, the local averaging is maintained by a wavelet-GALERKIN projection method as described in (7.1)–(7.3).
- 3 **Regularity.** By the SOBOLEV embedding given in Proposition B.1, we obtain the inclusions

$$W_2^2(\Omega \times T) \subset W_2^{2+\epsilon}(\Omega \times T) \subset C^{1/2}(\Omega \times T), \quad \epsilon > 0,$$

since $(\Omega \times T) \subset \mathbb{R}^3$. That is, choosing a finite wavelet approximation as in (5.13) with a representing wavelet frame of HÖLDER smoothness $s \geq 1/2$ (cf. Theorem 4.6 and Proposition 4.15), one is assured that

$$\vec{s}_0(\vec{x}, t) = (u_0, v_0, l_0, c_0)^T(\vec{x}, t) \in W_2^{2,(4)}(\Omega \times T)$$

holds true for the initial guess, as desired.

- 4 **Partial derivatives.** Under usage of optimized connection coefficients for differentiation (e.g. as in Example 7 of Chapter 6), the partial derivatives I_{x_1} , I_{x_2} and I_t are computed.
- 5 **Discretization.** In the next step, we have to discretize the system (9.6) in a suited way. Following the ideas developed in Chapter 7, a *semi-implicit* discretization is a good choice. This accelerates the numerical execution by allowing the feasible evolution steps to take a higher order of magnitude, see below. Using finite differences, the semi-implicit single-scale discretized version of (9.6) reads

$$\begin{aligned} u_{i,j}^{(k+1)} = & u_{i,j}^{(k)} - \tau \cdot \left(\alpha \cdot g_{i,j,(w)} \cdot \left(u_{i+1,j}^{(k+1)} - u_{i-1,j}^{(k+1)} + u_{i,j+1}^{(k+1)} - u_{i,j-1}^{(k+1)} \right) \right. \\ & \left. - I_{i,j,(x_1)} \cdot \left(I_{i,j,(x_1)} \cdot u_{i,j}^{(k+1)} + I_{i,j,(x_2)} \cdot v_{i,j}^{(k)} + I_{i,j,(t)} - I_{i,j} \cdot l_{i,j}^{(k)} - c_{i,j}^{(k)} \right) \right) \end{aligned} \quad (9.13)$$

$$\begin{aligned}
v_{i,j}^{(k+1)} &= v_{i,j}^{(k)} - \tau \cdot \left(\alpha \cdot g_{i,j,(w)} \cdot \left(v_{i+1,j}^{(k+1)} - v_{i-1,j}^{(k+1)} + v_{i,j+1}^{(k+1)} - v_{i,j-1}^{(k+1)} \right) \right. \\
&\quad \left. - I_{i,j,(x_2)} \cdot \left(I_{i,j,(x_1)} \cdot u_{i,j}^{(k)} + I_{i,j,(x_2)} \cdot v_{i,j}^{(k+1)} + I_{i,j,(t)} - I_{i,j} \cdot l_{i,j}^{(k)} - c_{i,j}^{(k)} \right) \right)
\end{aligned} \tag{9.14}$$

$$\begin{aligned}
l_{i,j}^{(k+1)} &= l_{i,j}^{(k)} - \tau \cdot \left(\beta \cdot g_{i,j,(l)} \cdot \left(l_{i+1,j}^{(k+1)} - l_{i-1,j}^{(k+1)} + l_{i,j+1}^{(k+1)} - l_{i,j-1}^{(k+1)} \right) \right. \\
&\quad \left. + I_{i,j} \cdot \left(I_{i,j,(x_1)} \cdot u_{i,j}^{(k)} + I_{i,j,(x_2)} \cdot v_{i,j}^{(k)} + I_{i,j,(t)} - I_{i,j} \cdot l_{i,j}^{(k)} - c_{i,j}^{(k)} \right) \right)
\end{aligned} \tag{9.15}$$

$$\begin{aligned}
c_{i,j}^{(k+1)} &= c_{i,j}^{(k)} - \tau \cdot \left(\gamma \cdot g_{i,j,(c)} \cdot \left(c_{i+1,j}^{(k+1)} - c_{i-1,j}^{(k+1)} + c_{i,j+1}^{(k+1)} - c_{i,j-1}^{(k+1)} \right) \right. \\
&\quad \left. + I_{i,j,(x_1)} \cdot u_{i,j}^{(k)} + I_{i,j,(x_2)} \cdot v_{i,j}^{(k)} + I_{i,j,(t)} - I_{i,j} \cdot l_{i,j}^{(k)} - c_{i,j}^{(k)} \right)
\end{aligned} \tag{9.16}$$

where $g_{i,j,(\cdot)}$ denotes the discretized diffusivity function $g(|\nabla \cdot|^2)$, all other notations should be self-explaining. The reader should note that we took advantage of a semi-implicit discretization of u and v also in the reaction part to gain more stability. Due to the different sign, this is prevented for l and c . We also point out at this stage, that it of course would be also possible and is in fact sensible to give a multiscale wavelet-based discretization of (9.6) under usage of the connection coefficients framework, since this would gain improved computational efficiency, see below.³

6 Evolutionary stability. In order to assure that an evolutionary nonlinear diffusion process remains stable, there are a set of criteria to be satisfied

³The multiscale scheme is more complicated in notation, but not in its computational execution; for the interested reader, we give it here in non-standard form for the parameter u , the expressions for the other three evolution equations follow in an equivalent manner.

$$\begin{aligned}
u_{i,j;\sigma;\tau}^{(k+1)} &= u_{i,j;\sigma;\tau}^{(k)} - \tau \cdot \left(\sum_{p=1}^{|q|} \sum_{l,m \in \Omega} \alpha \cdot g_{i,j;\sigma,(w)} \right. \\
&\quad \cdot \left(\left(\Gamma_{\varphi^{(p)}(\tau),[0,1]\tau}(\sigma, l, m, \sigma, i, j, 0) + \Gamma_{\varphi^{(p)}(\tau),[1,0]\tau}(\sigma, l, m, \sigma, i, j, 0) \right) \cdot u_{i,j;\sigma;\tau}^{(k+1)} \right. \\
&\quad \left. - I_{i,j;\sigma,(x_1)} \cdot \Gamma_{\varphi^{(p)}(\tau),[0,0]\tau}(\sigma, l, m, \sigma, i, j, 0) \right. \\
&\quad \left. \cdot \left(I_{i,j,(x_1)} \cdot u_{i,j;\sigma;\tau}^{(k+1)} + I_{i,j,(x_2)} \cdot v_{i,j;\sigma;\tau}^{(k)} + I_{i,j,(t)} - I_{i,j;\sigma;\tau} \cdot l_{i,j}^{(k)} - c_{i,j;\sigma;\tau}^{(k)} \right) \right)
\end{aligned}$$

Hereby, i, j are the spatial coordinates, σ denotes the scale of consideration and τ is an index pointing to the scaling function or wavelet $\varphi^{(\tau)}$. The notation for the connection coefficients is according to the conventions made in Chapter 6.

[235]. However, it is demonstrated in [237] that regularized nonlinear diffusion processes of PERONA-MALIK-type are unconditionally stable if a semi-discrete numerical scheme is applied. Unfortunately, the additional linear reaction term cancels this nice property for the proposed optical flow model and we have to choose τ in an appropriate way to guarantee the numerical stability of the system (9.13)–(9.16). A heuristic analysis of a linearized approximation to (9.6) justifies the choice $\tau = 1/2$. Additionally, we shall remind the reader that the choice of the optimal stopping time τ_s for the evolution is a purely experimental matter of a similar character as the determination of the weighting parameters α , β and γ . We choose a threshold ϵ and repeat the evolution until the norm of the systems residue understeps ϵ .

- 7 **Operator Splitting.** We will make use of the additive operator splitting as presented in Chapter 7 and extended in Section 8.8.3 to gain more computational efficiency for the implementation. First, we will rewrite the discrete diffusion-reaction equations (9.13)–(9.16) into the form

$$\begin{aligned}\vec{u}^{(k+1)} &= (\mathbf{Id} + \tau \cdot \mathbf{A}_u)^{-1} \cdot (\vec{u}^{(k)} - \tau \cdot \vec{b}_u) \\ \vec{v}^{(k+1)} &= (\mathbf{Id} + \tau \cdot \mathbf{A}_v)^{-1} \cdot (\vec{v}^{(k)} - \tau \cdot \vec{b}_v) \\ \vec{l}^{(k+1)} &= (\mathbf{Id} + \tau \cdot \mathbf{A}_l)^{-1} \cdot (\vec{l}^{(k)} - \tau \cdot \vec{b}_l) \\ \vec{c}^{(k+1)} &= (\mathbf{Id} + \tau \cdot \mathbf{A}_c)^{-1} \cdot (\vec{c}^{(k)} - \tau \cdot \vec{b}_c),\end{aligned}\tag{9.17}$$

where the entries of $\mathbf{A}_{(\cdot)}$ and $\vec{b}_{(\cdot)}$ are directly evident from (9.13)–(9.16). Now, the additive operator splitting can be directly applied to each of the equations in (9.17) in order to execute the necessary inversions. If we have chosen a multiscale discretization of (9.13)–(9.16) as proposed before, the application of the wavelet-extended additive operator splitting as described in Section 8.8.3 is also applicable in this situation leading to faster inversions for the predict part.

In order to obtain reliable flow estimations even for flow fields of large displacements, the whole method is embedded into a *coarse-to-fine* multiscale strategy. This means the following: before the flow estimation itself starts, the image sequence gets decomposed by a discrete wavelet transformation with a certain number of scales, depending on the image size. Then, the optical flow is computed — pursuant to the abovementioned processing steps — only on the lowpass part of the coarsest scale. The obtained result is backpropagated to the next finer scale by an inverse wavelet transform and rescaled accordingly. This new version is taken to be the *initial guess* for the flow field on the actual scale and the whole procedure starts from the beginning leading to successively improved estimations on stepwise refined scales. The process stops when the finest scale, i.e. the pixel level, is reached.

9.4. Extended Optical Flow via Segmentation Methods

We have already remarked that the proposed flow model (9.6) is by far not complete in the sense that it is capable to deal with all possible kinds of visual *occurrences* in image sequences. For some of these incidents of singular nature, we will sketch possible extensions to our original model in the actual section. These extensions are mainly based on the idea of *segmentation* which has a long tradition in image processing problems of various types by itself.

An inevitable property of moving objects observed by a camera is that they occlude and de-occlude other objects which are positioned *behind* them with respect to the visual ray of the camera during their motion. Furthermore, it is clear that it is impossible to make any precise statements about motion in occluded regions due to the lack of visual information. But on the other hand, one is interested to have an accurate mathematical model for this situation in order to determine motion boundaries and to separate different motions that appear to happen on the same local area of the image plane. To do this, let us denote the occluded regions of the image plane Ω by $\Xi \subset \Omega$. Following the ideas of MUMFORD and SHAH [162], we consider the minimization of a variational functional of the type

$$\begin{aligned} V(\vec{s}) = & \int_{\Omega \setminus \Xi} F(u, v, l, c)^2 + \alpha \cdot (|\nabla u|^2 + |\nabla v|^2) d\vec{x} \\ & + \int_{\Omega} \beta \cdot h(|\nabla l|^2) + \gamma \cdot h(|\nabla c|^2) d\vec{x} + |\Xi| = \min!, \end{aligned} \quad (9.18)$$

with $h' = g$ and F is defined by (9.3). The functional $V(\vec{s})$ may be interpreted in the following sense. The first integral is the optical flow constraint on the regular part $\Omega \setminus \Xi$ of the image plane with an additional linear regularization term. Since the implicit modelling of motion boundaries by the introduction of the set Ξ , it is not necessary to apply a nonlinear regularization method to the flow field. The second integral is the nonlinear regularization of the illumination conditions, it is the same as in the model (9.6). The last term penalizes irregular boundary structures of the occlusion set as well as its spatial expansion. Unfortunately, there are several theoretical difficulties coming along with the usage of (9.18). A very fundamental problem is that one (usually) has no a priori knowledge about the set Ξ at ones disposal, that is, one does not even know, how $V(\vec{s})$ does exactly look like. Mathematically, the functional is non-convex, which may cause the minimization procedure to get trapped in irrelevant local minima and moreover, theoretical as well as experimental investigations indicated additional problems of more practical nature, such as restrictive geometric constraints on junction and corner recovery or likely over-segmentation of the data [163] [185]. The basic idea to overcome these problems is the introduction of an additional

function $\xi : \mathbb{R}^2 \longrightarrow [0, 1]$ that plays the role of an indicator for the set Ξ , that is $\xi(\vec{x}) \approx \chi_\Xi$ or in other words,

$$\xi(\vec{x}) \approx \begin{cases} 1 & \text{if } \vec{x} \in \Xi \\ 0 & \text{otherwise.} \end{cases}$$

By the mathematical framework of Γ -convergence for functionals [71] [6], one is able to show that the new functional

$$\begin{aligned} \tilde{V}(\vec{s}, \xi) = & \int_{\Omega} (1 - \xi)^2 \cdot \left(F(u, v, l, c)^2 + \alpha \cdot (|\nabla u|^2 + |\nabla v|^2) \right) \\ & + \beta \cdot h(|\nabla l|^2) + \gamma \cdot h(|\nabla c|^2) + \eta |\nabla \xi|^2 + (4\eta)^{-1} \cdot \xi^2 d\vec{x} \end{aligned} \quad (9.19)$$

with the new parameter η is an approximation in the Γ -sense converging to $V(\vec{s})$ as $\eta \rightarrow 0$. The meaning of the first terms is almost self-explaining, they model the optical flow on the ξ -approximation of the set Ξ and the global illumination. The last two terms force ξ and $|\nabla \xi|$ to be small as long as the flow field gradient is also small and this is in fact *almost* the behaviour, an indicator function should obey. In order to obtain feasible solutions, any minimizer of (9.19) must necessarily fulfill the evolution system of the corresponding EULER differential equations given by the reaction-diffusion system

$$\begin{aligned} u_\tau &= (1 - \xi^2) \cdot \left(\alpha \cdot \Delta_{\vec{x}} u - I_{x_1} \cdot (I_{x_1} \cdot u + I_{x_2} \cdot v + I_t - I \cdot l - c) \right) \\ v_\tau &= (1 - \xi^2) \cdot \left(\alpha \cdot \Delta_{\vec{x}} v - I_{x_2} \cdot (I_{x_1} \cdot u + I_{x_2} \cdot v + I_t - I \cdot l - c) \right) \\ l_\tau &= \beta \cdot \operatorname{div} \left(g(|\nabla_{\vec{x}} l|^2) \cdot \nabla_{\vec{x}} l \right) + I \cdot (I_{x_1} \cdot u + I_{x_2} \cdot v + I_t - I \cdot l - c) \\ c_\tau &= \gamma \cdot \operatorname{div} \left(g(|\nabla_{\vec{x}} c|^2) \cdot \nabla_{\vec{x}} c \right) + I_{x_1} \cdot u + I_{x_2} \cdot v + I_t - I \cdot l - c \\ \xi_\tau &= \eta \cdot \Delta_{\vec{x}} \xi - (4\eta)^{-1} \cdot \xi + \alpha \cdot (1 - \xi) \cdot (|\nabla_{\vec{x}} u|^2 + |\nabla_{\vec{x}} v|^2). \end{aligned} \quad (9.20)$$

Since $\tilde{V}(\vec{s})$ is convex in ξ and we already know that (9.6) leads to a well-posed process under certain regularity assumptions, (cf. Theorem 9.1), we can expect the same to be true for the system (9.20). Recalling our considerations of the previous paragraph, the numerical implementation of (9.20) is straightforward and not carried out here for the sake of shortness.

The idea of approximating indicator functions and to couple them with the evolutionary process itself is also feasible to build extended optical flow models to deal with multiple motions — we shall comment on this in the following. Suppose there appear two motions, \vec{w}_1 and \vec{w}_2 , in the same local area of the image plane. In order to separate these, one should assign two different indicator

functions ξ_1, ξ_2 to these motions. The according optical flow model can be written as

$$\begin{aligned} \tilde{V}(\vec{s}_1, \vec{s}_2, \xi) = & \int_{\Omega} (1 - \xi_1)^2 \cdot \left(F(u_1, v_1, l, c)^2 + \alpha \cdot (|\nabla u_1|^2 + |\nabla v_1|^2) \right) \\ & + (1 - \xi_2)^2 \cdot \left(F(u_2, v_2, l, c)^2 + \alpha \cdot (|\nabla u_2|^2 + |\nabla v_2|^2) \right) \\ & + \beta \cdot h(|\nabla l|^2) + \gamma \cdot h(|\nabla c|^2) + \eta \cdot (|\nabla \xi_1|^2 + |\nabla \xi_2|^2) \\ & + (4\eta)^{-1} \cdot (\xi_1^2 + \xi_2^2) d\vec{x}, \end{aligned} \quad (9.21)$$

in a Γ -convergent variational formulation, which is an evident extension of (9.19). Since the evolutionary EULER differential equations and the discretized numerical schemes are obtainable in a straightforward way, these are also not carried out here. But the reader should be warned that a *good* approximation to the original MUMFORD-SHAH formulation generally causes numerical problems since very small choices for the approximation parameter η require likewise small evolution steps τ to keep the stability of the whole process guaranteed. This fact makes such methods inferior compared to classical nonlinear formulations like the model (9.6) in many practical situations, because a very large number of iteration steps is usually needed to obtain good results.

We close the discussion about segmentational flow models by pointing out another image occurrence which can not be detected by this method (and also not by the methods presented before). Namely, this model is not capable to deal with multiple motions of translucent objects, a very active area of research in the last few years. Such transparent motions are very hard to access in the spatial domain since there exists no proper way to separate the image into its two simultaneously visible planes and consequently, one cannot estimate the according partial derivatives which would be necessary for reliable flow calculations. Therefore, all approaches to transparent motions done so far were executed in the frequency domain where the image separation problem is decidable [16]. However, since this work only addresses spatial approaches via wavelet methods, translucency problems must remain unconsidered here.

9.5. Experimental Results

The actual section will demonstrate the suitability of the optical flow methods presented in the running of this chapter by applying it to some typical test image sequences and by testing the chosen modelling under artificial *laboratory conditions*. The evaluation of the results we will obtain in this section will be done under usage of two standard error measures which will be introduced now.

Average angular error. The result of an optical flow estimation is a vector field that shall describe the motion of objects in the considered image scene. Typically, one is interested to measure the error of the motion *direction*, that is the angular deviation from the correct motion field. Since velocity must be viewed as orientation in space *and* time, it was suggested in [15] to measure errors as the angular variations from the correct orientation in space-time. Embedding our original notation from the previous sections into the three-dimensional space-time, let

$$\vec{w} = [u, v, 1]$$

be the extended representation of the estimated optical flow field and similarly, let \vec{w}_c denote the correct values. Then the *angular error* between the correct and the estimated flow vectors is given by

$$\theta_{\text{err}} = \arccos\left(\left\langle \frac{\vec{w}}{\|\vec{w}\|}, \frac{\vec{w}_c}{\|\vec{w}_c\|} \right\rangle\right). \quad (9.22)$$

Accordingly, the *average angular error* is given by the mean value of all angular errors over the whole image plane Ω .

Average magnitude error. The other entity of interest in optical flow estimation is of course the *magnitude*, i.e. the length of the correct velocity vector. In contrast to the direction, the flow magnitude does not depend on space-time orientation and is thus considered only in the twodimensional image plane. For an estimation $\vec{w} = [u, v]$ and a ground truth \vec{w}_c , the *magnitude error* between the estimated and the real flow field is given by

$$m_{\text{err}} = \left| \|\vec{w}\| - \|\vec{w}_c\| \right|, \quad (9.23)$$

where $\|\cdot\|$ shall denote the EUCLIDEAN length in 2D. Again, the *average magnitude error* is obtained by taking the mean value over the whole image plane.

9.5.1. Testing illumination changes

In this paragraph, we shall investigate, how the proposed optical flow model behaves in practice when it comes to estimate over image sequences containing (temporal discontinuous) variations of the illumination conditions. To do this, we will create sets of random dot patterns with constant motion and check the capabilities of the model when certain illumination patterns are introduced.

Experiment 1. We create a set of random dot pattern frames of size 256×256 such that the motion between two successive frames is given by a constant horizontal velocity of 2 pixel/frame. A typical example is given in Figure 9.2. In



Figure 9.2. Input data for experiment 1. A random dot pattern of size 256×256 with constant horizontal motion of 2 pixels/frame.

this first experiment, we do not introduce any illumination changes and just apply the model (9.6) and a *reduced* version not considering illumination changes to the given data. The results for the sequence shown above are visualized in Figure 9.3. To evaluate the results and to gain some idea about the asymptotic

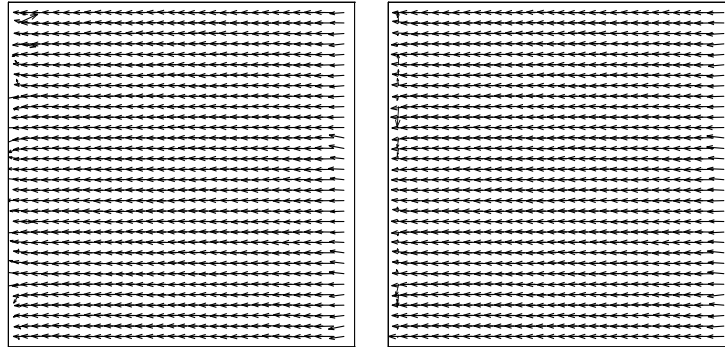


Figure 9.3. Results of experiment 1. *Left*. Flow estimation with the model (9.6) proposed in the previous section. *Right*. Flow estimation with the same model but without consideration of illumination changes.

quality of the applied methods, we repeated this experiment with 100 different sets of random dot sequences of size 256×256 . The averaged results in terms of the introduced error measures are collected in Table 9.1. One directly sees that both methods perform well and reach nearly the same estimation quality with respect to the average angular and magnitude errors. While they are visually more or less equivalent, regarding the numerical evaluation, the model not considering illumination changes performs even slightly better than the model proposed in (9.6). This may be explained by the fact that the least squares

	Av. ang. error	Av. mag. error
Flow model (9.6)	2.797^0	0.213 pixel
Model (9.6) without illumination	2.348^0	0.184 pixel

Table 9.1. Average performance results for experiment 1. Hereby, 100 random dot pattern sequences as in Figure 9.2 were used.

approximation in the prior case has a lower dimensionality, since there are two parameters less to be estimated (namely the illumination parameters l and c). Anyway, as we shall see in the following, the situation changes dramatically, as soon as new illumination conditions are introduced.

Experiment 2. The situation is very similar to that of the first experiment, but now, during the running of the sequence a constant additive illumination pattern is introduced in the right image half plane, see the illustration in Figure 9.4. The

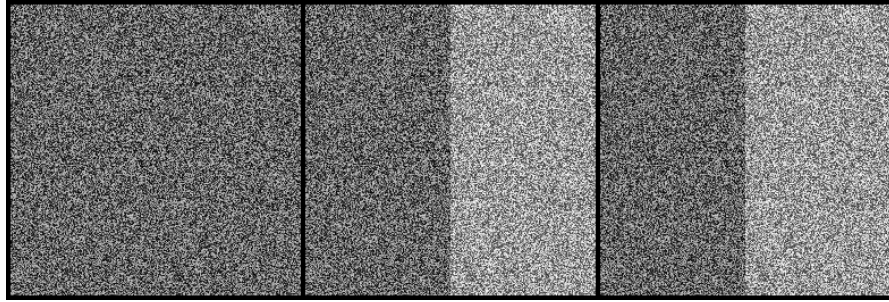


Figure 9.4. Input data for experiment 2. A random dot pattern of size 256×256 with constant horizontal motion of 2 pixels/frame where an additive illumination pattern occurs in the right half plane at a certain point of time.

estimation results are shown in Figure 9.5 and it is evident that the plain model is no longer capable to deal with this situation while the model (9.6) recovers the optical flow field almost correct. This is also confirmed by the evaluation results collected in Table 9.2. As in the first experiment, the results of 100 different random dot series were averaged to assure the reliability of the obtained values. Additionally, we show the estimated additive illumination field c in Figure 9.6.⁴

⁴Of course, applying the model (9.6) to the given data, the illumination would be measured in terms of both variables l and c . But to demonstrate how *additive* illumination changes are handled, we combined both parameters in such a way that only the additive part c is nontrivial.

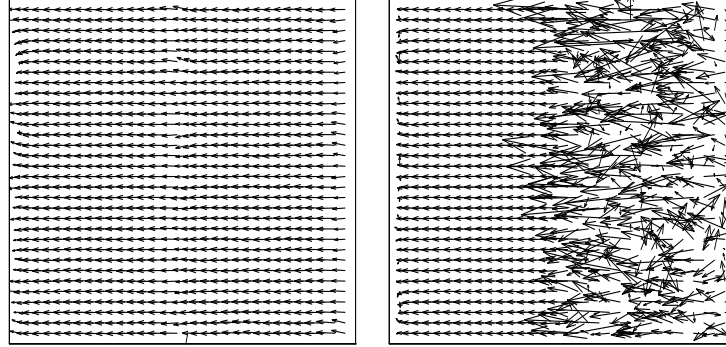


Figure 9.5. Results of experiment 2. *Left*. Flow estimation with the model (9.6) proposed in the previous section. *Right*. Flow estimation with the same model but without consideration of illumination changes.

Obviously, the introduced additive light source occurring in the right half plane is detected well, apart from some noise artefacts. However, these noise artefacts in the illumination pattern estimation do not distort the optical flow results.

	Av. ang. error	Av. mag. error
Flow model (9.6)	3.599 ^o	0.257 pixel
Model (9.6) without illumination	24.018 ^o	0.799 pixel

Table 9.2. Average performance results for experiment 2. Again, 100 random dot patterns of the kind as in Figure 9.4 were taken.

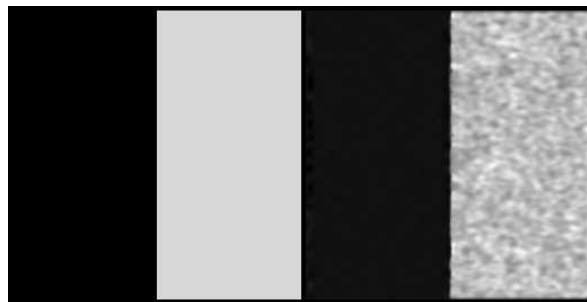


Figure 9.6. Comparison of illumination profiles for experiment 2. *Left*. Original profile. *Right*. Estimated additive illumination profile.

Experiment 3. Once again, we start with a sequence of random dot patterns with constant horizontal motion of 2 pixel/frame. This time, we introduce a

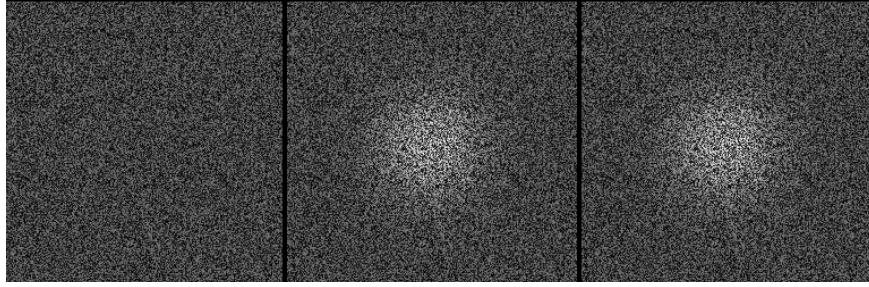


Figure 9.7. Input data for experiment 3. A random dot pattern of size 256×256 with constant horizontal motion of 2 pixels/frame where an multiplicative illumination pattern of GAUSSIAN shape (plus a constant) occurs at a certain point of time.

multiplicative illumination pattern of (up to an additive constant) GAUSSIAN shape in order to simulate the switching of new a light source. The resulting optical flow estimations for both methods are shown in Figure 9.7. As in the prior experiment, the plain model is not able to recover the optical flow in those image plane regions, where the illumination has a strong impact. In contrast to this, the model (9.6) gives satisfying results. This is confirmed by the evaluation of 100 repeated random experiments, see Table 9.3. To complete

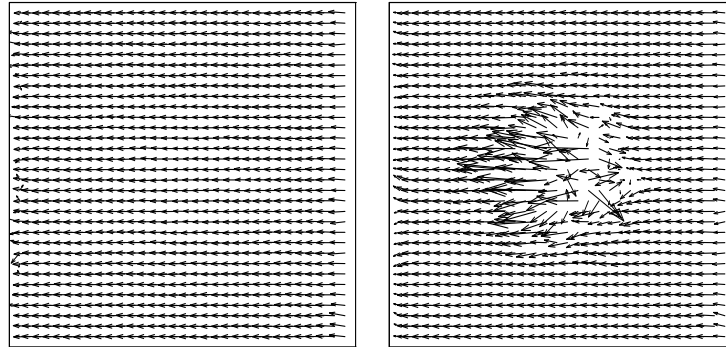


Figure 9.8. Results of experiment 3. *Left.* Flow estimation with the model (9.6) proposed in the previous section. *Right.* Flow estimation with the same model but without consideration of illumination changes.

this experiment, we give the estimated multiplicative illumination field $\lambda(\vec{x})$

	Av. ang. error	Av. mag. error
Flow model (9.6)	3.668 ⁰	0.272 pixel
Model (9.6) without illumination	8.728 ⁰	0.367 pixel

Table 9.3. Average performance results for experiment 3. Analogously to the first two experiments, 100 random dot patterns as in Figure 9.7 were applied.

in Figure 9.9.⁵ We notice that the illumination field $l(\vec{x})$ is also estimated in a satisfying manner up to possible boundary and noise artefacts which do not hinder reliable flow calculations.

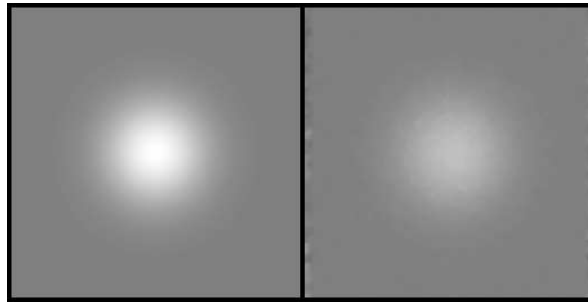


Figure 9.9. Comparison of illumination profiles for experiment 3. *Left.* Original profile. *Right.* Estimated multiplicative illumination profile.

Concluding these three experiments, modelling of possibly changing illumination conditions during the running of an image sequence can improve the flow estimations in the concerned image regions considerably. The obtained optical flow results are only slightly worse than those for image sequences having no illumination changes.

⁵In this experiment, we combined the illumination values l and c in such a way that only l is nontrivial, since we want to see the multiplicative illumination pattern. Since we have taken the shortcoming $l = (\mathbf{D}_t \lambda) / \lambda$ in (9.3), we have to re-estimate λ from l . This is done via the approximation

$$l = \frac{\mathbf{D}_t \lambda}{\lambda} \approx \frac{2}{\tau} \cdot \frac{\lambda(t + \tau) - \lambda(t)}{\lambda(t + \tau) + \lambda(t)}$$

and assuming that $\lambda(t) = 1$ since the light source is just activated during the time step from t to $t + \tau$. The resulting relation may be easily solved for $\lambda(t + \tau)$ yielding the desired approximation of the original illumination profile λ .

9.5.2. Recovery of motion boundaries

The following paragraph is dedicated to demonstrate the capabilities of the optical flow model (9.6) in the context of motion boundary recovery. As in the previous paragraph, the model is *tested* with a set of artificially created random patterns with known ground truth data.

The scenario is as follows: The dot pattern in the upper half of the considered image plane moves with constant speed of 1 pixel/frame from left to right and the pattern in the lower half moves contrarily. In Figure 9.10, the real and the estimated optical flow are shown. The estimation quality is of the same order of magnitude as in the first experiment. Since we want to test the ability of our implementation to recover flow boundaries in a proper way, we focus a little portion of the image plane and compare the results to that achieved by the *plain* model (9.3), this is visualized in Figure 9.11. The reduction of the smoothing effect introduced by the presmoothing and linear filtering is directly evident — the visual quality of the estimation from the nonlinear flow model (9.6) is significantly better than for the simple linear model (9.3), even though the boundary recovery is not perfect (but this can not be expected for such strongly discontinuous data like a random dot pattern).

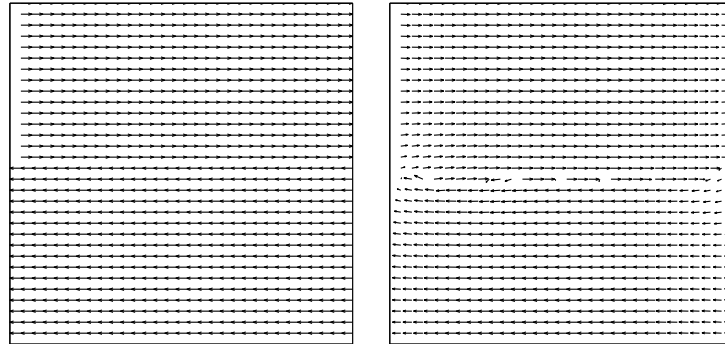


Figure 9.10. Results of experiment 4. *Left*. Real flow. *Right*. Flow estimation with the flow model (9.6).

9.5.3. Standard test images

To complete the experimental part of our optical flow investigations, we shall test the proposed model under some *more realistic* conditions by applying it to some standard test image sequences. For some of the test sequences, a set of ground truth data is available making explicit comparisons of the achieved

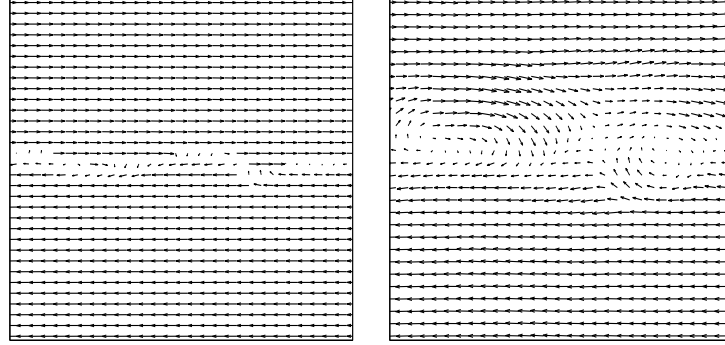


Figure 9.11. Focus of motion boundary. *Left.* Flow estimation with the model (9.6) proposed in the previous section. *Right.* Flow estimation with the simple model (9.3) and least squares approximation.

results possible.

The *Office* sequence. The first image sequence to be considered is the *Office* sequence created by GALVIN and co-workers [90], it is taken from the research website [250]. It shows a desk with a PC, a lamp, a picture and a window in the background and a chair in the foreground (see Figure 9.12). The occurring



Figure 9.12. First and last frame of the *Office* test sequence.

motion is a camera zoom-in where the focal point lies in the center of the image. The optical flow estimated with the model (9.6) is depicted in Figure 9.13, the average angular deviation from the real optical flow is evaluated to be 6.20° by means of (9.22). In comparison to the achieved results, the real optical flow

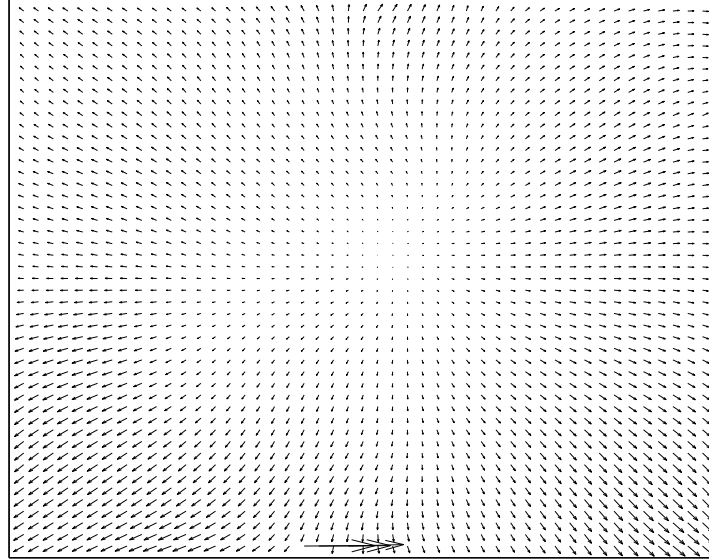


Figure 9.13. Estimated optical flow for the *Office* test sequence.

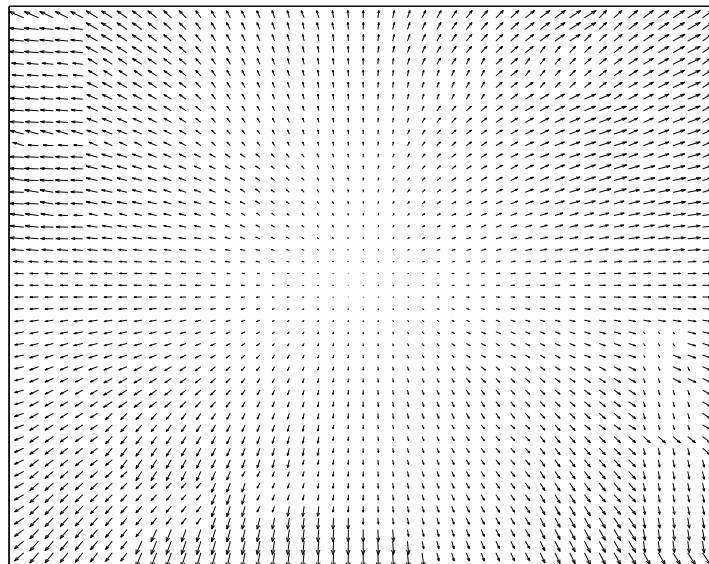


Figure 9.14. Real optical flow for the *Office* test sequence.

field for the *Office* scene is shown in Figure 9.14. Except to a very few numbers of significant misestimations in the lower middle of the image plane, the real and the estimated flow fields are visually nearly equivalent, which confirms the very low average angular error. Additionally, the average magnitude error is below 0.108 pixel, which is also a very good value.

The *Street* sequence. This example shows an artificially created road scene with a person sitting on a bank in the foreground and some houses plus a cloudy sky in the background. While a sport car is driving from the left to the right through the scene, the camera moves in the vice versa direction, this is shown in Figure 9.15. As mentioned above, this scene is created artificially



Figure 9.15. First and last frame of the *Street* test sequence.

and the graphic designer has also created a set of ground truth motion data that is (as the scene itself) available via [250]. The optical flow estimation achieved by our method (9.6) is shown in Figure 9.16, while the real flow data is visualized in Figure 9.17. One sees that the camera motion is recovered very well, but there are some estimation mistakes occurring below the front area of the moving car. Anyway, the overall performance is quite well, this can be seen in the comparison carried out in Table 9.4. We also point out that the average magnitude error is about 0.187 pixel. One final remark to the spatio-temporal approach of WEICKERT and SCHNÖRR: This idea is surely a theoretical step ahead and leads indeed to even better flow estimation results as is evident from Table 9.4, but on the other hand it is not a sensible presumption that the whole sequence is available at once, since this only allows flow estimation *in the past*.

The RUBICS *cube* scene. This is a very classical test sequence taken from the web resource [251]. It shows a RUBICS magic cube on a rotating plate and

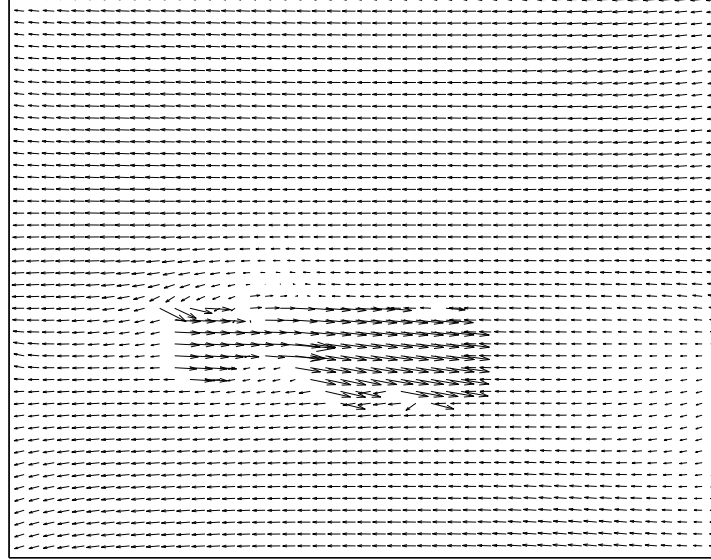


Figure 9.16. Estimated optical flow for the *Street* test sequence.

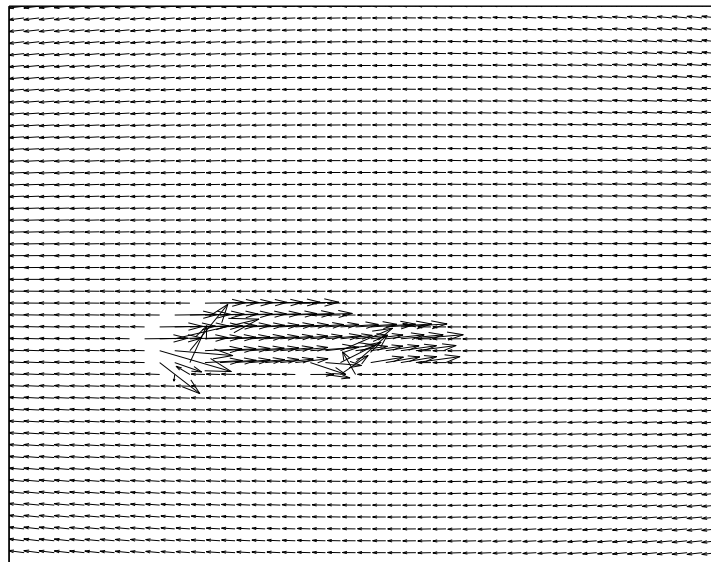


Figure 9.17. Real optical flow for the *Street* test sequence.

Author(s)	Av. ang. err.	Density
LUCAS & KANADE	$\approx 5^0$	32%
PROESMANS et. al.	$\approx 7^0$	100%
WEICKERT & SCHNÖRR	6.62^0	100%
WEICKERT & SCHNÖRR (spatio-temporal)	4.85^0	100%
K.N.	6.26^0	100%

Table 9.4. Various flow results for the *Street* sequence.

two still objects in the background. Two frames of this sequence are shown in Figure 9.18. Since this is a natural scene recorded by a camera, there exists no exact ground truth flow data and consequently, no exact assessment of the

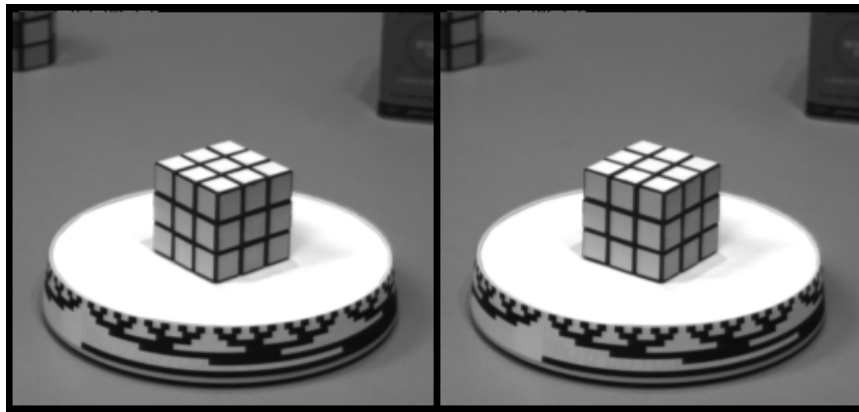


Figure 9.18. First and last frame of the RUBICS cube sequence.

flow results achieved by our algorithm (shown in Figure 9.19) is available. But the visual impression of the rotating plate and cube is reproduced well by the estimated optical flow vector field.

The Yosemite scene. Probably the most famous test sequence for optical flow algorithms is the *Yosemite* sequence that shows a computer graphically reproduced model of the Yosemite Valley and a virtual flight into the scene. Additionally, the clouds in the background move from left to right and the light intensity of the sky increases in some areas simulating sunlight effects. The starting and the ending frame of this sequence can be seen in Figure 9.20. As the previous example, this sequence is also taken from [251], which is also the source for the real optical flow data. The flow field estimated via (9.6)

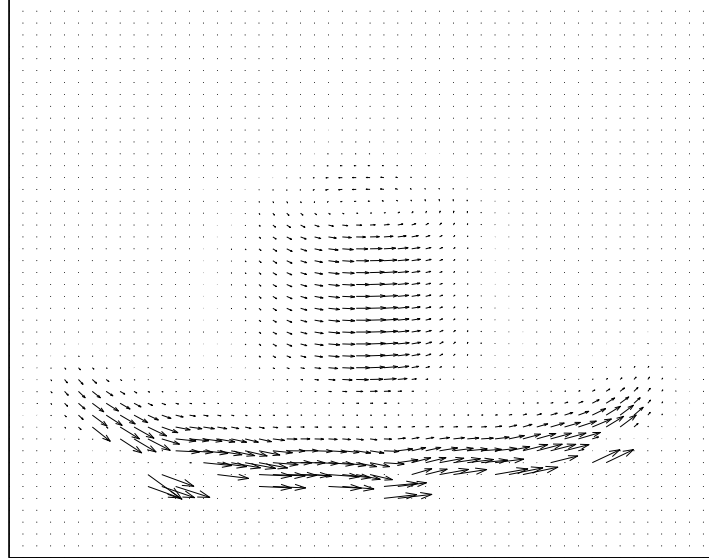


Figure 9.19. Optical flow estimated for the RUBICS *cube* sequence.

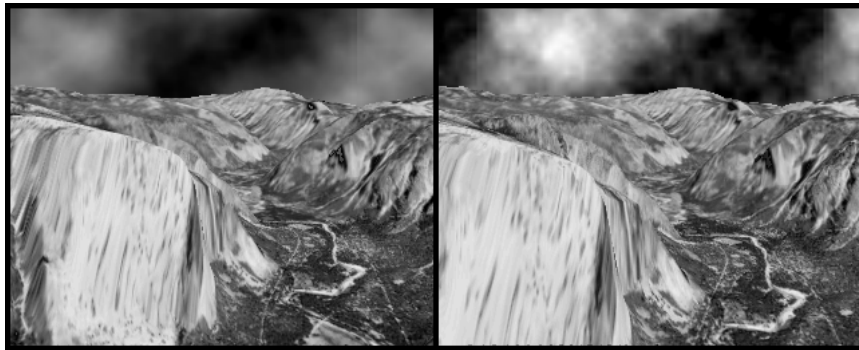


Figure 9.20. First and last frame of the *Yosemite* sequence.

is shown in Figure 9.21, while the correct flow field can be found in Figure 9.22. Both vector fields are visually very close to each other, except that the magnitude of the vectors of the correct flow field is usually slightly bigger than for those of the estimated flow field. This fact can be traced back to the nonlinear diffusion process, which has the property to equalize differences over a longer iteration period (see also the remarks in Chapter 8). But apart from this difference, the results for the estimation are once again very good in comparison

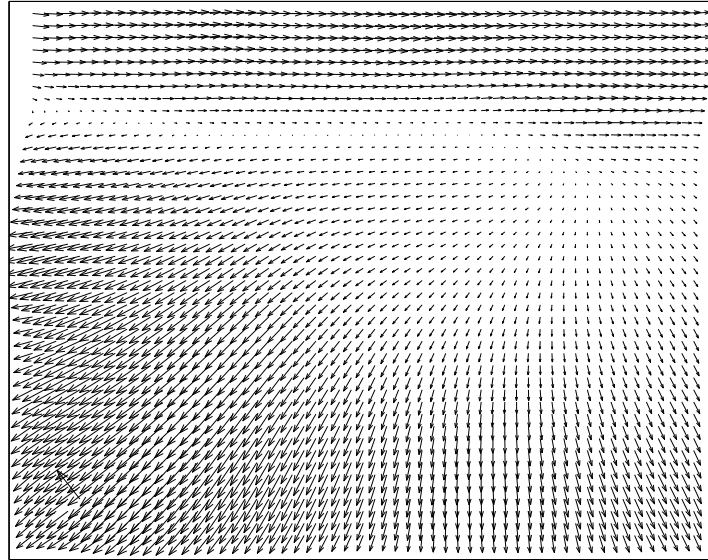


Figure 9.21. Estimated optical flow for the *Yosemite* test sequence.

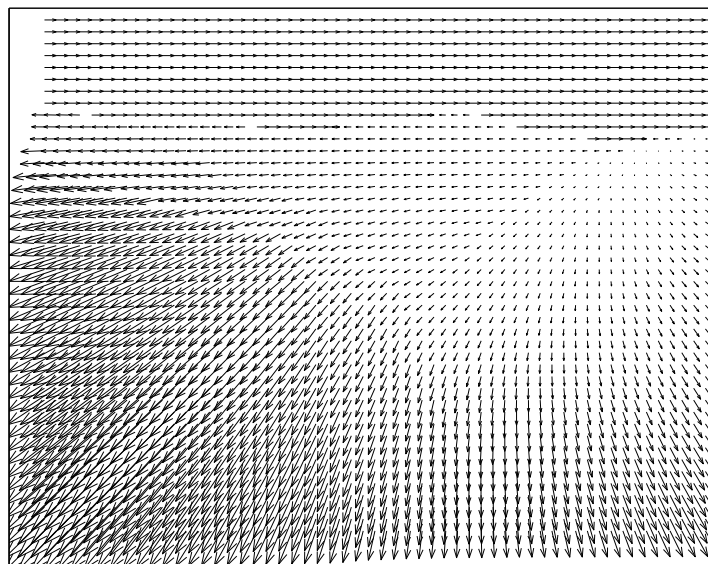


Figure 9.22. Real optical flow for the *Yosemite* test sequence.

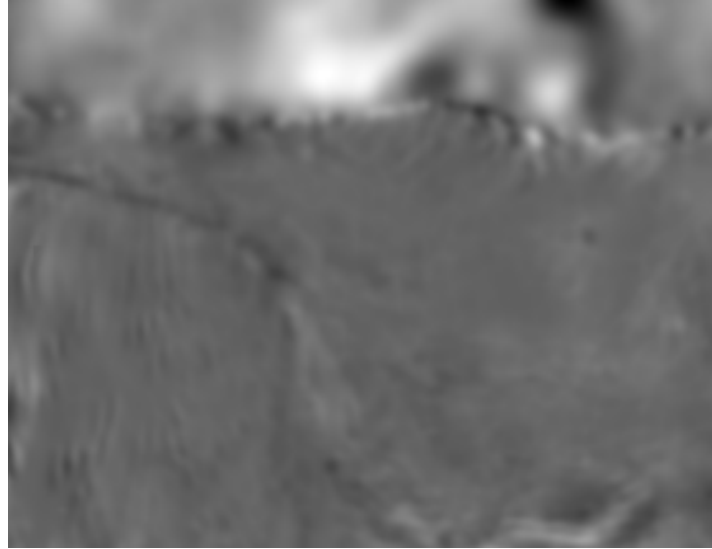


Figure 9.23. Estimated illumination profile for the *Yosemite* sequence.

to results achieved by other methods, see Table 9.5. To get an impression of the illumination change happening during the sequence, we have also depicted the estimated illumination profile (converted into a purely additive pattern) in Figure 9.23.

Author(s)	Av. ang. err.	Density	Method
FLEET & JEPSON	4.36 ⁰	34.1%	phase-based
WEBER & MALIK	4.31 ⁰	64.2%	local/differential
SINGH	12.90 ⁰	97.8%	region-based
URAS	10.44 ⁰	100%	local/differential
ALVAREZ ET AL.	5.53 ⁰	100%	global/differential
NAGEL	11.71 ⁰	100%	global/differential
K.N.	6.24 ⁰	100%	local/differential

Table 9.5. Various flow results for the *Yosemite* sequence.

In a last little experiment we have evaluated the impact of the proposed extensions from the simple flow model (9.3) to the extended model (9.6) as well as the influence of nonlinear wavelet estimation as suggested in Chapter

8. During the experiment, all steering parameters were left unchanged in order to assure equal conditions. The obtained qualities of the estimated optical flow fields are compared in Table 9.6 — they lead to the conclusion that the proposed extensions are in fact sensible, since they lead to distinct improvements.

<i>Yosemite</i> sequence	Av. ang. error
Model (9.3) without nonlinear wavelet postprocessing	7.59°
Model (9.6) without nonlinear wavelet postprocessing	6.76°
Model (9.3) with nonlinear wavelet postprocessing	6.53°
Model (9.6) with nonlinear wavelet postprocessing	6.24°

Table 9.6. Impact of proposed model extensions. Note how the nonlinear diffusion as well as the nonlinear wavelet estimation both improve the flow calculation.

9.6. Discussion and Related Work

To close the chapter about optical flow, we shall point out some research work that is related to our investigations. Of course, the differential approach to optical flow estimations under usage of the assumption that the image brightness does not change significantly over time leading to a formulation as in (9.2) is classical and fundamental for many solutions to the problem. It goes back to the pioneering work of HORN and SCHUNCK [114]. Just as old is the idea of regularizing this model in order to circumvent the underdetermination and the ill-posedness this formulation incorporates. While [114] suggested to use a linear regularization term, the idea of nonlinear flow regularization was first introduced in [165] and successively improved e.g. in [24], [238] and [4]. These latter two approaches are similar in spirit to the ideas developed in this work, since they also have their fundamentals in the idea of PERONA-MALIK regularization.

The illumination models proposed in Section 9.2 mainly go back to ideas of NEGAHDARIPOUR [173] [174] and his co-workers, some of these ideas were also proposed in the thesis [19]. But all these approaches only considered the improved modelling of illumination changes but did not regard the simultaneously available possibilities of nonlinear regularization or vice versa. By combining the capabilities of both illumination modelling and nonlinear regularization strategies, we could obtain improved estimation results for optical flow fields. This is especially the case if illumination changes as well as distinct motion boundaries occur in the considered image sequences, see our experiments with sets of random dot patterns. If we apply our method to the standard test image sequences, which only have slight illumination changes and

moderate boundary structures, the performance is about the same as for other well-performing standard methods. This should be no too big surprise, since these latter ones are modelled in a similar way than ours, see above. Anyway, even in this situation, our approach has the advantage to incorporate the usage of wavelet-optimized operator representations (Chapter 6) and to be highly efficient thanks to the reduced computational amount, if the wavelet additive operator splitting is applied.

Chapter Summary

This chapter was dedicated to the problem of computing the optical flow occuring in a sequence of images. After reviewing the fundamental problems and the main paradigms how to solve this problem, we have introduced our point of view, namely the differential geometric approach. Since the plain model turned out to be inadequate in many typical situations, the model was extended such that it also took illumination changes into account. Additionally, we have introduced a nonlinear regularization process in order to guarantee the sharp recovery of motion boundaries and to compensate for the spatial overlapping coming along with the (by all means necessary) linear prefiltering. In the next section, we have proven that the proposed model is indeed consistent and leads to a well-posed process. We have also pointed out its concrete implementation by describing the method stepwise and demonstrating that the required prerequisites are indeed met. Furthermore, we have sketched some more possible model extensions basing on the idea of segmentation. In this context, the meaning of the MUMFORD-SHAH functional for image processing purposes was enlightened. The chapter was closed by a large experimental part that showed the various capabilities of the proposed method. It performed similarly well to the best optical flow methods known so far.

Chapter 10

CONCLUSION

And if I can't do this, it will be done by those who come after me

—LEOPOLD KRONECKER

The research presented during the running of this thesis was mainly concerned with the development of construction methods for nonseparable wavelets in higher dimensions, with the provision of the according mathematical analytical methods and with the concrete implementation of these results in the practical context of computer vision applications. This closing chapter shall provide a brief review over the achieved results and also provide a vague outlook into possible future directions.

10.1. Summary

In this section, the main results presented in this thesis shall be summarized. Hereby, particular emphasis is put on the original contributions of the author and several possible consequences of the achieved results are commented.

In Chapter 1, we gave some motivation for the present research work by reviewing the development of wavelet theory from its roots in the beginning of the 20th century till today. To enable the reader to comprehend the motivation to apply wavelet theory to computer vision problems, we also gave an introduction to the fundamental concepts and paradigms of computer vision. The chapter was closed by an overview, how the thesis is organized.

The second chapter was dedicated to introduce the basic mathematical concepts of time-frequency decompositions and onedimensional wavelet theory to equip the reader with the basic terminology. In regard to problems that occur in

multidimensional wavelet filter design, we introduced a new filter construction method that does not depend on factorization theorems like the FEJER-RIESZ Theorem. This new method was based on the concept of (linear independent) basic filters, the main results are found in Lemma 2.13 and Proposition 2.15. The working way of this concept was demonstrated by a short example.

Chapter 3 was dedicated to the investigation of dilation matrices in multiple dimensions. Here, we investigated the induced properties like the geometric shape of the incorporated sampling lattice and we proposed a quality measure to assess, how close a multidimensional sampling scheme can be to ideal (i.e. radial symmetric) sampling. Furthermore, we established a general design principle for dilation matrices and gave a classification for the special class of dyadic regular sampling (Proposition 3.7). This chapter was closed by a survey about the asymptotic behaviour for several classes of dilation matrices under increasing dimensionality. The according results are collected in Theorem 3.8 and the following corollaries.

The first part of the thesis was closed by Chapter 4. Here, the main point of interest was the difficult problem of wavelet (or scaling) filter design in higher dimensions. Since factorization approaches were no longer feasible, we focussed ourselves to direct design methods and to dimensional reduction methods. This dimensional reduction, established in Lemma 4.2 and Lemma 4.3, provided us with some knowledge about how *subfilters* of (yet unknown) multidimensional filters must look like. This technique turned out to reduce the complexity of the filter design leading to some new higher order orthogonal wavelet filters (see also the examples in Appendix A). After a discussion of some general design principles and some excursions to concrete feature implementations, some generalizations of orthogonal wavelets were presented. While biorthogonal wavelets were only briefly reviewed, the (for the applications we have in mind) very attractive concept of tight wavelet frames was presented in much more detail leading to a new and very simple construction method (Theorem 4.13 and Lemma 4.14) for such function families. The investigations about tight wavelet frames were closed by a statement about the asymptotic regularity of wavelets built via this method (Proposition 4.15).

In Chapter 5, the fundamentals of multidimensional wavelet analysis were collected. After reviewing some classical results about the projection operators associated to a wavelet basis and the multidimensional variant of the MALLAT algorithm, some approximation results are given. Some of them had to be slightly weakened to fit into the multidimensional framework (this was the case for Proposition 5.3 and for a following corollary).

The main chapter of the second part was Chapter 6. In this chapter, the efficient representation of various operators in bases of nonseparable, multidimensional wavelets was developed. Here, we generalized the concept of *connection coefficients* to the multidimensional case and derived a way to calculate these coefficients via simple linear equation systems (main result is Theorem 6.7). The working way of this framework was demonstrated at hand of a standard differential operator; an example and a method to optimize operators to fit to certain requirements were also given here. In the following, the efficient representation of affine transforms in wavelet coordinated was carried out, since these play a prominent role in several vision models. The chapter was closed by a review of different ways to put these operator representation into matrix form. This is a reproduction of results obtained by BEYLKIN et. al. [20], but again generalized to the multidimensional formulation.

Chapter 7 recalls various ways to put operator equations of differential type into a form that is suited for numerical processing. Hereby, GALERKIN approximations, multiscale preconditioning and various types of operator decompositions are presented. All these reformulations are stated in the multidimensional wavelet formalism.

The application oriented part of this thesis started with Chapter 8. Linear and nonlinear scale spaces were considered here. We introduced a new linear wavelet scale space by slightly modifying the corresponding axiomatics. This linear wavelet scale space helped us to reduce the computational amount for nonlinear regularization via additive operator splitting by *shrinking* the size of the operator matrix that has to be inverted. Additionally, we have presented a new type of nonlinear wavelet scale space that is based on the approximation in BESOV spaces, which is a direct consequence of the results from [41].

The final Chapter 9 dealt with the computation of optical flow. After motivating the approach, several model additions to handle illumination changes and motion boundaries were introduced. A detailed investigation of the numerical behaviour was carried out (Theorem 9.1) and several hints for the concrete implementation were also detailed. In an extensive experimental part, the model is tested in regard of the proposed properties. It turned out that the proposed optical flow model behaves well for the considered test image sequences, while it should be even faster than most of the other methods in an optimized wavelet based implementation that takes advantage of wavelet additive operator splitting as given at the end of Chapter 8.

To close this summary, we should also point out some remarks to the aspect of nonseparability, which was one of the original motivations for this research.

The achieved improvements in comparison to separable methods turned out to be of lower significance in the most numerical experiments. For this reason, we have not put an according juxtaposition into the experimental part. Anyway, there were slight improvements noticeable in most experiments and the reader is also reminded of the optimization of discrete differentiation schemes in Chapter 6, where the nonseparability also yielded improvements. And despite from the aspect of numerical and/or computational improvements, one should always keep in mind that the nonseparable framework is the much more general and flexible mathematical concept that already contains the separable case as a subset. Even for this reason only, the usage of this more universal techniques seems to be sensible.

10.2. Outlook

This thesis shall be closed by a vague look into the future. Many possible future directions were already sketched during the running of the text at the according places. Now, we should point explicitly on some of the more urgent and especially desirable extensions to be realized.

There hasn't been much research done in the area of nonseparable, multidimensional wavelet theory so far. And the results achieved are only a first little step in that direction, many questions still remain open. For example, it is still unknown, whether there exist orthogonal wavelet families of increasing order in higher dimensions as a generalization of the DAUBECHIES wavelets. The answer to this questions would rely on the solution of certain polynomial factorization results — as soon as these are accessible (which might happen tomorrow or never), the existence problem would be solved. Another theoretical aspect of interest is the paraunitary matrix completion in order to find associated finite highpass filters for a non-dyadic lowpass scaling filter. Some special results in this direction were achieved by LAWTON and co-workers, but the general case is still unclear. Problems of this kind do no longer occur for tight wavelet frames; the results by RON and SHEN and those developed in this thesis make wavelet frames of arbitrary order accessible.

From a more practical point of view, betterments in the numerical application as well as in the implementation are a challenging future task. For example, the introduction of novel discretization schemes, that could make use of sparsity patterns in the wavelet-represented data sets, might lead to efficiency and/or stability improvements. In the same context, we should also mention the lack of software implementations of nonseparable wavelet algorithms — the *classical* tool packages like the MatLab[®] wavelet toolbox or the free WaveLab do not provide code for nonseparable wavelet transformations. The development of appropriate tools is strongly desired in order to improve nonseparable algo-

rithms and to stimulate a greater number of researchers to work with these.

The implementation of nearly all up-to-date optical flow algorithms does not fit through the bottleneck of being employed in real-time applications. One might wait another five or six years till the available computers will be fast enough to be capable of this. More interesting would be the parallel implementation of current algorithms, since most of them — especially the wavelet based ones as the model proposed in this thesis — have an excellent potential to be realized on parallel machines. Furthermore, general improvements of the existing flow models are also desirable, e.g. it might be of interest to combine the advantages of spatial and spectral models to solve problems like translucency in a more satisfying way. Another critical point with existing models is the choice of the free steering parameters. Some general rules for the (more or less) automatic determination of *optimal* parameter choices are also desirable with regard to the application of optical flow algorithms in autonomous machines.

*So eine Arbeit wird eigentlich nie fertig,
man muß sie für fertig erklären,
wenn man nach Zeit und Umständen
das Möglichste getan hat.*

—JOHANN WOLFGANG VON GOETHE

Appendix A

A Little Dictionary Of Wavelet Filters

Research is what I'm doing when I don't know what I'm doing.

—WERNHER VON BRAUN

In this appendix, we present a collection of some wavelet filters obtained under usage of the methods developed in the first part of this thesis. In the upcoming part of this appendix, we will give a number of orthogonal dyadic scaling filters, first for the onedimensional case, then for the twodimensional quincunx case. Orthogonal filters play only a minor role with respect to the applications this thesis addresses, but they are important for themselves as a mathematical *prototype* for the set of *all* wavelet filters. This shall be understood in the way that the design problem for any tight wavelet frame filter is essentially the same as for orthogonal ones, only the degree of allowed filters and unknown coefficients differs. The reader may compare the fundamental relations (4.1) and (4.19) to see this. Consequently, all the fundamental mathematical problems occurring in the design of discrete wavelet filters may be studied and understood in the orthogonal case. The second part of this appendix gives two examples of twodimensional wavelet frames obtained by the generalized filter design methods for tight wavelet frames. These frames are mainly used implicitly in the computer vision applications presented in the third part of this work.

Orthogonal Filters

1D filters. DAUBECHIES orthogonal scaling filters derived from the basic filter concept developed in Section 2.2. Notation of the basic filters corresponds to (2.20) and n indicates the degree of the wavelet.

- $n = 1$ (HAAR scaling filter)

$$h = \frac{1}{2} \cdot g^{0,1}$$

- $n = 2$ (Classical DAUBECHIES, tap-4)

$$h = \frac{1}{8} \cdot g^{0,3} + \frac{\sqrt{3}}{8} \cdot g^{1,2}$$

- $n = 3$ (DAUBECHIES, tap-6)

$$h = \frac{1}{32} \cdot g^{0,5} + \frac{\sqrt{5+2\sqrt{10}}}{32} \cdot g^{1,4} + \frac{\sqrt{10}}{32} \cdot g^{2,3}$$

- $n = 4$ (DAUBECHIES, tap-8)

$$h = \frac{1}{128} \cdot g^{0,7} + \frac{\lambda}{128} \cdot g^{1,6} + \frac{\lambda^2 - 7}{256} \cdot g^{2,5} + \frac{\sqrt{35}}{128} \cdot g^{3,4}$$

with

$$\lambda = \frac{1}{\sqrt{3}} \left(v_3 \pm \sqrt{-v_1^2 v_2 + \frac{2}{\sqrt{7}} v_1 v_3 + \frac{1}{\sqrt{7}} v_1^2 v_3 + \frac{2\sqrt{21}}{\sqrt{5}} - \frac{\sqrt{3}}{\sqrt{35}} v_1^2 v_3 - \frac{4\sqrt{3}}{\sqrt{35}} v_1 v_3 - v_1 + 14} \right)$$

and the abbreviations

$$\begin{aligned} v_1 &= \sqrt[3]{154 + 42\sqrt{15}} \\ v_2 &= \frac{11 - 3\sqrt{15}}{14} \\ v_3 &= \sqrt{v_1 + 7 + v_1^2 v_2}. \end{aligned}$$

- $n = 5$ (DAUBECHIES, tap-10)

$$h = \frac{1}{512} \cdot g^{0,9} + \frac{\lambda}{512} \cdot g^{1,8} + \frac{\lambda^2 - 9}{1024} \cdot g^{2,7} + \frac{\lambda^4 - 18\lambda^2 - 63 + 24\sqrt{14}}{4096\lambda} \cdot g^{3,6} + \frac{\sqrt{126}}{512} \cdot g^{4,5}$$

with

$$\lambda = \left(9 + 2v_2 \pm 2 \left(\frac{v_1^2 v_2}{100} + 24 + 12\sqrt{14} - v_1 + \frac{16v_2}{5} + \frac{\sqrt{14}v_1 v_2}{20} - \frac{v_1^2 v_2}{300} + \frac{7i\sqrt{15}v_1^2 v_2}{300} + \frac{12\sqrt{14}v_2}{5} - \frac{3v_1 v_2}{5} - \frac{i\sqrt{210}v_1^2 v_2}{300} - \frac{3i\sqrt{210}v_1 v_2}{100} + \frac{v_1^2}{16} + \frac{7i\sqrt{15}v_1^2}{240} + \frac{7i\sqrt{15}v_1 v_2}{75} \right)^{\frac{1}{2}} \right)^{\frac{1}{2}}$$

and the abbreviations

$$\begin{aligned} v_1 &= \sqrt[3]{-225 + 105i\sqrt{15}} \\ v_2 &= \sqrt{6\sqrt{14} + 12 - \frac{7i\sqrt{15}}{240} v_1^2 - \frac{1}{16} v_1^2 + v_1}. \end{aligned}$$

Such closed form descriptions were conjectured to be inaccessible for $n > 3$ by several authors (e.g. [184]), but under application of the basic filter framework, at least the cases $n = 4, 5$ are still feasible as seen above. Similar closed form descriptions can also be found for onedimensional coiflet filters up till filter length 12.

2D quincunx filters. These can be viewed as the twodimensional analogues to the DAUBECHIES filters; they were obtained using Lemma 4.2. Again, n denotes the degree of the filters.

- $n = 1$ (2D-HAAR scaling filter, same as in 1D)

$$h = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

- $n = 2$ (KOVAČEVIĆ-VETTERLI scaling filter [127])

$$h = \begin{bmatrix} \frac{-1-\sqrt{3}}{16} & \frac{\frac{3+3\sqrt{3}}{16}}{\frac{3+\sqrt{3}}{16}} & \frac{\frac{3+\sqrt{3}}{16}}{\frac{3-\sqrt{3}}{16}} & \frac{-1+\sqrt{3}}{16} \end{bmatrix}$$

- $n = 3$ (New higher order scaling filter consisting of 20 taps — inaccessible before. Filter is given componentwise to fit on page)

$$\begin{aligned} h_{3,4} &= \frac{31 - 5v_1 - 9v_2 + 5v_3 + 4v_1v_2 - v_1v_3 + 7v_2v_3 - 2v_1v_2v_3}{256} \\ h_{4,4} &= \frac{-5 + 11v_1 - 27v_2 + 15v_3 + 12v_1v_2 - 3v_1v_3 - 5v_2v_3 + 2v_1v_2v_3}{256} \\ h_{5,4} &= \frac{-255 + 87v_1 + 125v_2 - 45v_3 - 38v_1v_2 + 15v_1v_3 - 75v_2v_3 + 24v_1v_2v_3}{256} \\ h_{6,4} &= \frac{85 - 29v_1 + 143v_2 - 55v_3 - 23v_1v_2 + 17v_1v_3 + 25v_2v_3 - 8v_1v_2v_3}{256} \\ h_{2,3} &= \frac{-57 + 24v_1 + 57v_2 - 21v_3 - 15v_1v_2 + 6v_1v_3 - 17v_2v_3 + 5v_1v_2v_3}{384} \\ h_{3,3} &= \frac{135 - 18v_1 + 99v_2 - 35v_3 - 27v_1v_2 + 8v_1v_3 + 27v_2v_3 - 9v_1v_2v_3}{384} \\ h_{4,3} &= \frac{2025 - 621v_1 - 1395v_2 + 555v_3 + 450v_1v_2 - 177v_1v_3 + 541v_2v_3 - 172v_1v_2v_3}{768} \\ h_{5,3} &= \frac{-1275 + 399v_1 - 1497v_2 + 585v_3 + 474v_1v_2 - 183v_1v_3 - 391v_2v_3 + 124v_1v_2v_3}{768} \\ h_{6,3} &= \frac{-135 + 27v_1 + 45v_2 - 25v_3 - 18v_1v_2 + 7v_1v_3 - 39v_2v_3 + 12v_1v_2v_3}{768} \\ h_{7,3} &= \frac{45 - 9v_1 + 63v_2 - 27v_3 - 18v_1v_2 + 9v_1v_3 + 13v_2v_3 - 4v_1v_2v_3}{768} \\ h_{1,2} &= \frac{45 - 9v_1 - 63v_2 + 27v_3 + 18v_1v_2 - 9v_1v_3 + 13v_2v_3 - 4v_1v_2v_3}{768} \\ h_{2,2} &= \frac{-135 + 27v_1 - 45v_2 + 25v_3 + 18v_1v_2 - 7v_1v_3 - 39v_2v_3 + 12v_1v_2v_3}{768} \\ h_{3,2} &= \frac{-1275 + 399v_1 + 1497v_2 - 585v_3 - 474v_1v_2 + 183v_1v_3 - 391v_2v_3 + 124v_1v_2v_3}{768} \\ h_{4,2} &= \frac{2025 - 621v_1 + 1395v_2 - 555v_3 - 450v_1v_2 + 177v_1v_3 + 541v_2v_3 - 172v_1v_2v_3}{768} \\ h_{5,2} &= \frac{135 - 18v_1 + 99v_2 + 35v_3 + 27v_1v_2 - 8v_1v_3 + 27v_2v_3 - 9v_1v_2v_3}{384} \\ h_{6,2} &= \frac{-57 + 24v_1 - 57v_2 + 21v_3 + 15v_1v_2 - 6v_1v_3 - 17v_2v_3 + 5v_1v_2v_3}{384} \\ h_{2,1} &= \frac{85 - 29v_1 - 143v_2 + 55v_3 + 23v_1v_2 - 17v_1v_3 + 25v_2v_3 - 8v_1v_2v_3}{256} \\ h_{3,1} &= \frac{-255 + 87v_1 - 125v_2 + 45v_3 + 38v_1v_2 - 15v_1v_3 - 75v_2v_3 + 24v_1v_2v_3}{256} \\ h_{4,1} &= \frac{-5 + 11v_1 + 27v_2 - 15v_3 - 12v_1v_2 + 3v_1v_3 - 5v_2v_3 + 2v_1v_2v_3}{256} \\ h_{5,1} &= \frac{31 - 5v_1 + 9v_2 - 5v_3 - 4v_1v_2 + v_1v_3 + 7v_2v_3 - 2v_1v_2v_3}{256} \end{aligned}$$

with the shortcomings

$$\begin{aligned}\gamma_1 &= \sqrt{10} \\ \gamma_2 &= \sqrt{5 + 2\sqrt{10}} \\ \gamma_3 &= \sqrt{17 + 8\sqrt{10}}.\end{aligned}$$

Hereby, the filter coefficients are arranged as follows

$$\mathbf{h} = \begin{bmatrix} & & \mathbf{h}_{3,4} & \mathbf{h}_{4,4} & \mathbf{h}_{5,4} & \mathbf{h}_{6,4} & \\ & \mathbf{h}_{2,3} & \mathbf{h}_{3,3} & \mathbf{h}_{4,3} & \mathbf{h}_{5,3} & \mathbf{h}_{6,3} & \mathbf{h}_{7,3} \\ \mathbf{h}_{1,2} & \mathbf{h}_{2,2} & \mathbf{h}_{3,2} & \mathbf{h}_{4,2} & \mathbf{h}_{5,2} & \mathbf{h}_{6,2} & \\ & \mathbf{h}_{2,1} & \mathbf{h}_{3,1} & \mathbf{h}_{4,1} & \mathbf{h}_{5,1} & & \end{bmatrix}.$$

This latter example is the first algebraically given third order quincunx scaling filter published so far (at least to the authors best knowledge). Such filters are usually very hard to obtain due to the high computational complexity of the design procedure. The complexity reduction could be achieved under application of Lemma 4.2 and the conjunction of various design principles. Again, it is possible to derive coiflet filters of similar order, see e.g. Example 4 in Chapter 4.

Filters For Tight Wavelet Frames

In this section, we will present two filter sets leading to tight wavelet frames with frame bound 1. These examples shall illustrate the capabilities of the mechanisms developed in the section about multidimensional filter design possibilities for tight wavelet frames.

2D tight frame for regular sampling matrix. Constructed via the subfilter method of Lemma 4.3. All wavelets have three vanishing moments and are in \mathcal{C}^2 . This frame is implicitly used in all the experiments executed in the third part of this work.

$$\begin{aligned}\mathbf{h} &= \frac{1}{64} \begin{bmatrix} 1 & 3 & 3 & 1 \\ 3 & 9 & 9 & 3 \\ 3 & 9 & 9 & 3 \\ 1 & 3 & 3 & 1 \end{bmatrix} \\ g^1 &= \frac{1}{128} \begin{bmatrix} 13 - 3\sqrt{5} & 3 + 3\sqrt{5} & -3 - 3\sqrt{5} & -13 + 3\sqrt{5} \\ 3 + 3\sqrt{5} & -3 - 3\sqrt{5} & 3 + 3\sqrt{5} & -3 - 3\sqrt{5} \\ 3 + 3\sqrt{5} & -3 - 3\sqrt{5} & 3 + 3\sqrt{5} & -3 - 3\sqrt{5} \\ 13 - 3\sqrt{5} & 3 + 3\sqrt{5} & -3 - 3\sqrt{5} & -13 + 3\sqrt{5} \end{bmatrix} \\ g^2 &= \frac{1}{128} \begin{bmatrix} -13 + 3\sqrt{5} & -3 - 3\sqrt{5} & -3 - 3\sqrt{5} & -13 + 3\sqrt{5} \\ -3 - 3\sqrt{5} & 3 + 3\sqrt{5} & -3 - 3\sqrt{5} & 3 + 3\sqrt{5} \\ -3 - 3\sqrt{5} & 3 + 3\sqrt{5} & -3 - 3\sqrt{5} & 3 + 3\sqrt{5} \\ 13 - 3\sqrt{5} & 3 + 3\sqrt{5} & -3 - 3\sqrt{5} & -13 + 3\sqrt{5} \end{bmatrix} \\ g^3 &= \frac{1}{128} \begin{bmatrix} 0 & 0 & 3\sqrt{6} - \sqrt{30} & -3\sqrt{6} + \sqrt{30} \\ 0 & 0 & 13\sqrt{6} + \sqrt{30} & 3\sqrt{6} - \sqrt{30} \\ -3\sqrt{6} + \sqrt{30} & -13\sqrt{6} - \sqrt{30} & 0 & 0 \\ 3\sqrt{6} - \sqrt{30} & -3\sqrt{6} + \sqrt{30} & 0 & 0 \end{bmatrix}\end{aligned}$$

$$\begin{aligned}
g^4 &= \frac{1}{128} \begin{bmatrix} -3\sqrt{6} + \sqrt{30} & 3\sqrt{6} - \sqrt{30} & 0 & 0 \\ 3\sqrt{6} - \sqrt{30} & 13\sqrt{6} + \sqrt{30} & 0 & 0 \\ 0 & 0 & -13\sqrt{6} - \sqrt{30} & -3\sqrt{6} + \sqrt{30} \\ 0 & 0 & -3\sqrt{6} + \sqrt{30} & 3\sqrt{6} - \sqrt{30} \end{bmatrix} \\
g^5 &= \frac{1}{64} \begin{bmatrix} 4\sqrt{3} - \sqrt{15} & \sqrt{15} & \sqrt{15} & 4\sqrt{3} - \sqrt{15} \\ \sqrt{15} & -4\sqrt{3} - \sqrt{15} & -4\sqrt{3} - \sqrt{15} & \sqrt{15} \\ \sqrt{15} & -4\sqrt{3} - \sqrt{15} & -4\sqrt{3} - \sqrt{15} & \sqrt{15} \\ 4\sqrt{3} - \sqrt{15} & \sqrt{15} & \sqrt{15} & 4\sqrt{3} - \sqrt{15} \end{bmatrix} \\
g^6 &= \frac{1}{128} \begin{bmatrix} -3\sqrt{3} + \sqrt{15} & 3\sqrt{3} - \sqrt{15} & -3\sqrt{3} + \sqrt{15} & 3\sqrt{3} - \sqrt{15} \\ 3\sqrt{3} - \sqrt{15} & 13\sqrt{3} + \sqrt{15} & -13\sqrt{3} - \sqrt{15} & -3\sqrt{3} + \sqrt{15} \\ -3\sqrt{3} + \sqrt{15} & -13\sqrt{3} - \sqrt{15} & 13\sqrt{3} + \sqrt{15} & 3\sqrt{3} - \sqrt{15} \\ 3\sqrt{3} - \sqrt{15} & -3\sqrt{3} + \sqrt{15} & 3\sqrt{3} - \sqrt{15} & -3\sqrt{3} + \sqrt{15} \end{bmatrix} \\
g^7 &= \frac{1}{128} \begin{bmatrix} 13 - 3\sqrt{5} & 3 + 3\sqrt{5} & -3 - 3\sqrt{5} & -13 + 3\sqrt{5} \\ 3 + 3\sqrt{5} & -3 - 3\sqrt{5} & 3 + 3\sqrt{5} & -3 - 3\sqrt{5} \\ -3 - 3\sqrt{5} & 3 + 3\sqrt{5} & -3 - 3\sqrt{5} & 3 + 3\sqrt{5} \\ -13 + 3\sqrt{5} & -3 - 3\sqrt{5} & 3 + 3\sqrt{5} & 13 - 3\sqrt{5} \end{bmatrix}
\end{aligned}$$

The attentive reader may have registered that the lowpass filter of this frame is spatially separable. But this does not contradict the philosophy of this thesis, since the corresponding highpass filters (which eventually describe structures in the considered data) are all nonseparable and built in such a way that they guarantee a possibly close-to-isotropic processing. In this example, we chose a separable lowpass filter, because this gave the best regularity and isotropy properties for the corresponding scaling function.

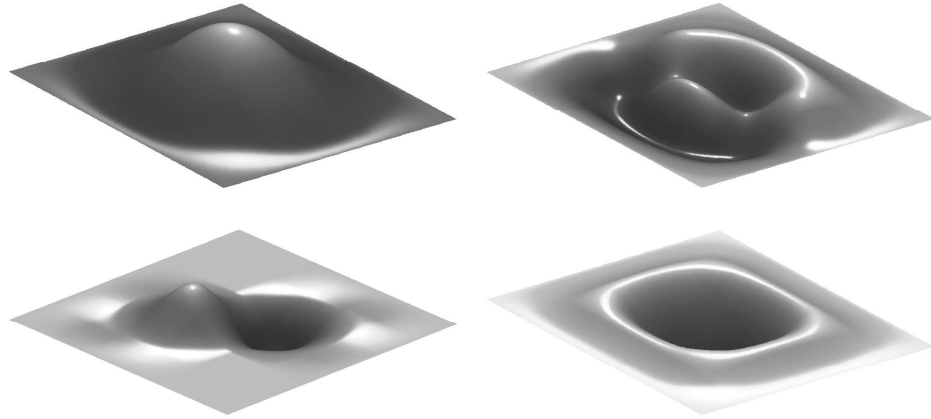


Figure A.1. Frame functions h , g^1 , g^3 and g^5 from the previous example.

Second example. To demonstrate the working way of Theorem 4.13 and Lemma 4.14, consider the (quincunx) dyadic orthogonal filter set

$$\begin{aligned} \mathbf{h} &= \begin{bmatrix} 0 & \frac{35}{58} & \frac{7}{29} \\ -\frac{3}{29} & \frac{15}{58} & 0 \end{bmatrix} \\ \mathbf{g} &= \begin{bmatrix} 0 & \frac{15}{58} & \frac{3}{29} \\ \frac{7}{29} & -\frac{35}{58} & 0 \end{bmatrix}. \end{aligned}$$

The filter coefficients were chosen in such a way that the transfer function of the scaling filter is as symmetric as possible. In order to get a (partially) symmetric tight wavelet frame from this orthogonal family, we can apply Lemma 4.14 yielding the new filters

$$\begin{aligned} \tilde{\mathbf{h}} &= \begin{bmatrix} 0 & 0 & \frac{525}{3364} & 0 & -\frac{21}{841} \\ 0 & \frac{100}{841} & \frac{1}{2} & \frac{100}{841} & 0 \\ -\frac{21}{841} & 0 & \frac{525}{3364} & 0 & 0 \end{bmatrix} \\ \tilde{\mathbf{g}}^1 &= \begin{bmatrix} 0 & 0 & -\frac{1225}{3364}\sqrt{2} & 0 & \frac{49}{841}\sqrt{2} \\ 0 & \frac{105}{841}\sqrt{2} & 0 & \frac{105}{841}\sqrt{2} & 0 \\ -\frac{9}{841}\sqrt{2} & 0 & \frac{225}{3364} & 0 & 0 \end{bmatrix} \\ \tilde{\mathbf{g}}^2 &= \begin{bmatrix} 0 & 0 & -\frac{525}{3364} & 0 & \frac{21}{841} \\ 0 & -\frac{100}{841} & \frac{1}{2} & -\frac{100}{841} & 0 \\ \frac{21}{841} & 0 & -\frac{525}{3364} & 0 & 0 \end{bmatrix}. \end{aligned}$$

As stated in Lemma 4.14, the scaling filter is obviously interpolating and this one as well as one of the highpass filters are indeed symmetric. Moreover, it can be shown that while the starting filter gave wavelets which are not even continuous, this new frame reaches a smoothness order of $s \approx 0.775 \dots$ in terms of the critical HÖLDER exponent.

Appendix B

Function Spaces

Although our last, not least!

—WILLIAM SHAKESPEARE [206]

In this appendix, the definitions and some basic facts about the function spaces used throughout this thesis are collected. We focus ourselves to those properties that are necessary to supplement the investigations within this thesis. For a complete presentation of function spaces theory, the reader may consider [221] [222].

The LEBESGUE spaces $L_p(\Omega)$. Let $p \geq 1$ be any real number and $\Omega \subset \mathbb{R}^n$. For all functions $f : \Omega \rightarrow \mathbb{R} \cup \{\pm\infty\}$, we introduce the pseudo-norm

$$\|f\|_p^* = \left(\int_{\mathbb{R}^n} |f(\vec{x})|^p d\vec{x} \right)^{1/p} \in \mathbb{R}_+ \cup \{\infty\}. \quad (\text{B.1})$$

It is an easy task to verify that all locally integrable functions f with $\|f\|_p^* < \infty$ form a vector space which will be denoted by \mathcal{V}_p^* . Then, the set

$$N = \{ f \in \mathcal{V}_p^* \mid \|f\|_p^* = 0 \}$$

consisting of all functions of zero measure, is a vector subspace of \mathcal{V}_p^* . Now, we define the quotient

$$L_p(\Omega) = \mathcal{V}_p^* / N. \quad (\text{B.2})$$

The pseudo-norm $\|f\|_p^*$ induces a norm on the space $L_p(\Omega)$ which will be denoted $\|f\|_{L_p}$. Equipped with this norm, $L_p(\Omega)$ becomes a BANACH space. For the special case $p = 2$, we define a scalar product

$$\langle f_1, f_2 \rangle = \int_{\Omega} f_1(\vec{x}) \cdot \overline{f_2(\vec{x})} d\vec{x},$$

which makes $L_2(\Omega)$ a HILBERT space with $\|f\|_{L_2}^2 = \langle f, f \rangle$. It is exactly this HILBERT space property involving several further features (e.g. the PARSEVAL relation) which makes the space L_2 the best suited fundamental space to do harmonic analysis and signal processing.

The HÖLDER spaces $\mathcal{C}^\alpha(\Omega)$. For $k \in \mathbb{N}$, $\mathcal{C}^k(\Omega)$ shall denote the well-known class of bounded, continuous functions f on $\Omega \subset \mathbb{R}^n$ having k continuous derivatives $\partial_{\vec{x}}^{\vec{j}} f$ with $|\vec{j}| \leq k$. This definition may be generalized to non-integer *smoothness exponents* $\alpha = k + s$ with $0 < s \leq 1$ in the following manner. Let

$$\omega(f, h) = \sup\{|f(\vec{x}_1) - f(\vec{x}_2)| : |\vec{x}_1 - \vec{x}_2| \leq h\} \quad (\text{B.3})$$

denote the *modulus of continuity*. The HÖLDER spaces $\mathcal{C}^\alpha(\Omega)$ are then defined by

$$\mathcal{C}^\alpha(\Omega) = \{f : \Omega \rightarrow \mathbb{R} : \omega(\partial_{\vec{x}}^{\vec{j}} f, h) \leq C \cdot h^s \ \forall \ |\vec{j}| \leq k\}. \quad (\text{B.4})$$

Endowed with the norm

$$\|f\|_{\mathcal{C}^{k+s}} = \sum_{0 \leq |\vec{j}| \leq k} \|\partial_{\vec{x}}^{\vec{j}} f\|_\infty + \sup_{h, |\vec{j}|=k} (\omega(\partial_{\vec{x}}^{\vec{j}} f, h) \cdot h^s), \quad (\text{B.5})$$

$\mathcal{C}^\alpha(\Omega)$ becomes a BANACH space. For $s = 0$, this definition is inadequate and we must replace the *classical* continuity classes by the ZYGMUND classes which are defined as above but use the following varied modulus of continuity

$$\tilde{\omega}(f, h) = \sup\{|f(\vec{x}_1 + \vec{x}_2) + f(\vec{x}_1 - \vec{x}_2) - 2 \cdot f(\vec{x}_1)| : |\vec{x}_1 - \vec{x}_2| \leq h\}.$$

In this thesis, HÖLDER spaces are only implicitly used to measure the smoothness of several wavelets designed by the developed methods. But there is in fact much more to do with them, for example it is possible to measure the HÖLDER smoothness of any function only by considering its wavelet coefficients [157] or to embed them into (for us more interesting) SOBOLEV spaces, see below.

The generalized SOBOLEV spaces $W_p^k(\Omega)$. In order to combine the capacities of LEBESGUE spaces and smoothness spaces, we define the generalized SOBOLEV spaces

$$W_p^k(\Omega) = \{f : \partial_{\vec{x}}^{\vec{j}} f \in L_p(\Omega) \ \forall \ 0 \leq |\vec{j}| \leq k\}. \quad (\text{B.6})$$

Sometimes, these are also called SOBOLEV-SLOBODECKIJ spaces in the mathematical literature. They can be supplied with the norm

$$\|f\|_{W_p^k} = \left(\int_\Omega \sum_{0 \leq |\vec{j}| \leq k} |\partial_{\vec{x}}^{\vec{j}} f(\mathbf{x})|^p d\mathbf{x} \right)^{1/p}. \quad (\text{B.7})$$

An equivalent norm is given by

$$\|f\|_{W_p^s} = \left\| (1 + |\vec{\omega}|^2)^{s/2} \cdot \hat{f}(\vec{\omega}) \right\|_{L_p}, \quad (\text{B.8})$$

which makes the generalization to non-integer smoothness exponents as well as the extension to negative smoothness exponents straightforward; both issues are not detailed any further here.

PROPOSITION B.1 *The following inclusions are valid:*

- 1 $W_p^{s+t}(\Omega) \subset W_p^s(\Omega)$ for all $s \geq 0$ and $t \geq 0$.
- 2 $W_p^s(\Omega) \subset \mathcal{C}^k(\Omega)$ for $s > n/p + k$ and $k \in \mathbb{N}_0$.
- 3 $W_\infty^{s+1+\epsilon}(\Omega) \subset W_1^s(\Omega) \subset \mathcal{C}^s(\Omega)$ for $\epsilon > 0$.

4 $f \in \mathcal{C}^s(\Omega) \Rightarrow f \in W_\infty^s(\Omega)$ if f is compactly supported.

The following lemma is very useful for solutions of partial differential equations or regularization of ill-posed problems.

LEMMA B.2 *The spaces $W_2^s(\Omega)$ and $W_2^{-s}(\Omega)$ are dual spaces associated to the standard L_2 inner product for all $s \in \mathbb{R}$.*

The BESOV spaces $B_{p,q}^s(\Omega)$. For $r \in \mathbb{N}$ and $\vec{j} \in \mathbb{R}^n$ we define the r -th difference of a function f as

$$d_{\vec{j}}^r f(\vec{x}) = \sum_{k=0}^r \binom{r}{k} \cdot (-1)^k \cdot f(\vec{x} + k\vec{j}),$$

which is defined on the set

$$\Omega(r, \vec{j}) = \{ \vec{x} \in \Omega : \vec{x} + r \cdot \vec{j} \in \Omega \}.$$

Generalizing (B.3), we set the r -th $L_p(\Omega)$ modulus of smoothness for $0 < p \leq \infty$ to

$$\omega(f, h, r, p) = \sup_{|\vec{j}| \leq h} \|d_{\vec{j}}^r f(\vec{x})\|_{L_p(\Omega(r, \vec{j}))}. \quad (\text{B.9})$$

Given $s > 0$ and $1 \leq p, q \leq \infty$, choose $r \in \mathbb{N}$ such that $r - 1 \leq s < r$. Then, the BESOV space $B_{p,q}^s(\Omega)$ can be defined by the norm

$$\|f\|_{B_{p,q}^s} = \|f\|_{L_p} + \left(\int_0^\infty (h^{-s} \cdot \omega(f, h, r, p))^q \frac{dh}{h} \right)^{1/q}. \quad (\text{B.10})$$

This technical definition can be understood in the following way: $B_{p,q}^s(\Omega)$ is the space of L_p functions having s derivatives such that the finite differences of f for each scale may be controlled by l_q sequences. As before, the definition can be extended such that the smoothness exponent s can also attain negative values. BESOV spaces can be related to SOBOLEV and HÖLDER spaces in the following way.

LEMMA B.3 *Let $1 \leq p < \infty$. Then $W_p^s(\Omega) = B_{p,p}^s(\Omega)$ and $\mathcal{C}^s(\Omega) = B_{\infty,\infty}^s(\Omega)$ hold true.*

PROPOSITION B.4 ([41]) *Given a tight wavelet frame Ψ for $L_2(\Omega)$ built from a dilation matrix \mathbf{Q} , having at least $m > s$ vanishing moments such that $\psi \in B_{p,q}^t(\Omega)$ for all $\psi \in \Psi$ and $t > s$, we have the following norm equivalence measured in terms of the wavelet coefficients $c_k^{j,\psi} = \int_\Omega \psi_{j,\vec{k}} \cdot f$*

$$\|f\|_{B_{p,q}^s} \simeq \left(|c^{0,\varphi}|^{q/p} + \sum_j \left(\sum_{\vec{k}, \psi} |\det \mathbf{Q}|^{j \cdot (p-2) + j \cdot s \cdot p} \cdot |c_k^{j,\psi}|^p \right)^{q/p} \right)^{1/q}. \quad (\text{B.11})$$

It is an easy task to show that $B_{p,p}^s(\Omega)$ is embedded in $L_2(\Omega)$ for $s \cdot p + p - 2 \geq 0$. Obviously, the characterization (B.11) becomes especially simple for the spaces with $p = q$, since this decouples the norm estimates within the scales. Another case of special interest in practice is $p = s = 1$; the reason will be given in Lemma B.5. BESOV spaces can be further generalized to the wavelet oscillation spaces $O_p^{s_1, s_2}$ introduced by JAFFARD for detailed singularity investigations within the wavelet domain [115]; these spaces will not be considered here.

All the spaces we have presented so far have one thing in common: they may be completely characterized by wavelet coefficients as in (B.11) and this property makes these spaces especially comfortable to work with them in the wavelet domain. But on the other hand, for applications in computer vision, other function spaces, which do not share this pleasant property, also play an important role. A particular space of this kind will be defined now.

The space $BV(\Omega)$. Given any function $f \in L_1(\Omega)$, its *variation* is defined by

$$V_\Omega(f) = \int_\Omega |\nabla f(\vec{x})| \, d\vec{x}, \quad (\text{B.12})$$

where the gradient operator is understood in a distributional sense. The space of *bounded variation* $BV(\Omega)$ is thus the space of all $L_1(\Omega)$ -functions such that $V_\Omega(f) < \infty$. Since (B.12) only induces a semi-norm, $BV(\Omega)$ becomes a BANACH space endowed with the norm

$$\|f\|_{BV(\Omega)} = \|f\|_{L_1(\Omega)} + V_\Omega(f). \quad (\text{B.13})$$

An issue of practical interest in image processing is, whether a function space contains non-continuous functions to e.g. recover edges in images and simultaneously smoothen in areas of lower variation. While the *nice* BESOV spaces of type $B_{1,q}^1(\Omega)$ do not have this property, such functions are admissible in $BV(\Omega)$. However, $BV(\Omega)$ is at least *close* to BESOV spaces in the following sense.

LEMMA B.5 $B_{1,1}^1(\Omega) \subset BV(\Omega) \subset B_{1,\infty}^1(\Omega)$.

References

The library is the mathematician's laboratory.

—PAUL R. HALMOS [104]

- [1] AKANSU, ALI N. , HADDAD, RICHARD A. and CAGLAR, HAKAN. *The Binomial QMF-Wavelet Transform for Multiresolution Signal Decomposition*. IEEE Trans. on Signal Processing, 41(1), 13–19. 1993.
- [2] ALPERT, B., BEYLKIN, G., GINES, D. and VOZOVoi, L.. *Adaptive Solution of Partial Differential Equations in Multiwavelet Bases*. PAM Report 409, submitted. 1999.
- [3] ALVAREZ, L., GUICHARD, F., LIONS, P.-L. and MOREL, J.-M. *Axioms and Fundamental Equations in Image Processing*. Arch. Rational Mech. Analysis, 123, 199–257. 1993.
- [4] ALVAREZ, L., ESCLARÍN, J., LEFÉBURE, M. and SÁNCHEZ, J. *A PDE Model for Computing the Optical Flow*. Proceedings XVI Congreso de Ecuaciones Diferenciales y Aplicaciones. Las Palmas de Gran Canaria — España. 1999.
- [5] ALVAREZ, LUIS and WEICKERT, JOACHIM and SÁNCHEZ, JAVIER. *Reliable Estimation of Dense Optical Flow Fields with Large Displacements*. International Journal of Computer Vision. 2000.
- [6] AMBROSIO, L. and TORTORELLI, V. *Approximation of Functionals Depending on Jumps by Elliptic Functionals via Γ -convergence*. Boll. Un. Mat. Ital., 7, 105–123. 1992.
- [7] ANANDAN, P. *A Computational Framework and an Algorithm for the Measurement of Visual Motion*. Int. Journal Computer Vision, 2, 283–310. 1992.
- [8] ANTONINI, M., BARLAUD, M., MATHIEU, P. and DAUBECHIES, I. *Image Coding Using Wavelet Transforms*. IEEE Trans. Image on Processing, 1, 205–220. 1992.
- [9] AUBERT, G., DERICHE, R. and KORNPBST, P. *A Mathematical Study of the Regularized Optical Flow Problem in the Space $BV(\Omega)$* . Technical Report 503. Université de Nice-Sophia Antipolis 1997.
- [10] AUBERT, G. and KORNPBST, P. *A Mathematical Study of the Relaxed Optical Flow Problem in the Space $BV(\Omega)$* . SIAM Journal on Numerical Analysis. 1998.
- [11] AYACHE, ANTOINE. *Construction of Non Separable Dyadic Compactly Supported Orthonormal Wavelet Bases for $L^2(\mathbb{R}^2)$ of Arbitrarily Smoothness*. Revista Matematica Iberoamericana, 15(1), 37–58. 1999.

- [12] BAJSCY, R. *Active Perception*. Proceedings IEEE, 996–1005. 1988.
- [13] BALIAN R. *Un Principe d'Incertitude Fort en Théorie du Signal ou en Mécanique Quantique*. C. R. Acad. Sci. Paris, 292, Série II, 1357–1361. 1981.
- [14] BARANIUK, R. G. and CHOI, H. *Information-theoretic Analysis of Besov Spaces*. Proc. SPIE Tech. Conference on Wavelet Applications in Signal and Image Processing VIII. San Diego. 2000.
- [15] BARRON, J. L., FLEET, D. J. and BEAUCHEMIN, S. S. *Performance of Optical Flow Techniques*. Int. Journal on Computer Vision, 12, 43–77. 1994.
- [16] BEAUCHEMIN, S. S. and BARRON, J. L. *A Theory of Occlusion in the Context of Optical Flow*. Advances in Computer Vision, 191–200. SOLINA, F., KROPATSCH, W., KLETTE, R. and BAJCSY, R. (eds.). Springer Verlag, Wien — New-York — Heidelberg. 1997.
- [17] BELOGAY, EUGENE and WANG, YANG. *Arbitrarily Smooth Orthogonal Nonseparable Wavelets in \mathbb{R}^2* . SIAM J. Math. Anal. 1997.
- [18] BENEDETTO, JOHN J. and FRAZIER, MICHAEL W. (eds.). *Wavelets: Mathematics and Applications*. CRC Press, Boca Raton. 1994.
- [19] BERNARD, CHRISTOPHE. *Fast Optic Flow with Wavelets*. **In:** Proceedings Wavelets and Applications Workshop (WAW '98). 1998.
- [20] BEYLKIN, G., COIFMAN, R. and ROKHLIN, V. *Fast Wavelet Transforms and Numerical Algorithms, I*. Comm. Pure and Appl. Math., 44, 141–183. 1991.
- [21] BEYLKIN, G. *On the Representation of Operators in Bases of Compactly Supported Wavelets*. SIAM Journal on Numerical Analysis, 29(6), 1716–1740. 1991.
- [22] BEYLKIN, G. *Wavelets, Multiresolution Analysis and Fast Numerical Algorithms*. **In:** Wavelets — Theory and Applications [82], 182–262. 1996.
- [23] BEYLKIN, GREGORY and KEISER, JAMES M. *An Adaptive Pseudo-Wavelet Approach for Solving Nonlinear Partial Differential Equations*. **In:** Multiscale Wavelet Methods for Partial Differential Equations. W. DAHMEN, A. KURDILA and P. OSWALD (eds.). Wavelet Analysis and its Applications Vol. 6, 137–198. Academic Press, San Diego — London — Boston — New York — Sydney — Tokyo — Toronto. 1999.
- [24] BLACK, M. J. and ANANDAN, P. *Robust Dynamic Motion Estimation Over Time*. Proc. IEEE Conf. Computer Vision and Pattern Recognition, 296–302. 1991.
- [25] DE BOOR, C., HÖLLIG, K. and RIEMENSCHNEIDER, S. *Box Splines*. Applied Mathematical Sciences, 98. ISBN 0-387-94101-0. Springer Verlag, Berlin — Heidelberg — New York. 1993.
- [26] BRONSTEIN, I.N., SEMENDJAJEW, K.A., MUSIOL, G. and MÜHLIG, H. *Taschenbuch der Mathematik*. ISBN 3–8171–2004–4. Verlag Harri Deutsch, Frankfurt am Main. 1999.
- [27] BÜLOW, THOMAS. *Hypercomplex Spectral Signal Representations for Image Processing and Analysis*. PhD thesis. Christian-Albrechts-Universität zu Kiel, 1999.

- [28] BURT, P. J. *Fast Filter Transforms for Image Processing*. Computer Vision, Graphics and Image Processing, 16, 20–51. 1981.
- [29] BURT, P. and ADELSON, T. *The Laplacian Pyramid as a Compact Image Code*. IEEE Trans. Commun., 31, 532–540. 1983.
- [30] BYRNES, JIM (editor). *Twentieth Century Harmonic Analysis — A Celebration*. Proceedings of the NATO Advanced Study Institute. Mathematics, Physics and Chemistry, vol. 33. Kluwer Academic Publishers, Dordrecht — Boston — London. 2000.
- [31] CABRELLI, CARLOS and MOLTER, URSULA. *Generalized Self-Similarity*. Jour. Math. Anal. Appl., 230, 251–260. 1999.
- [32] CABRELLI, CARLOS, HEIL, CHRISTOPHER and MOLTER, URSULA. *Accuracy of Several Multidimensional Refinable Distributions*. Preprint, School of Mathematics, Georgia Institute of Technology. 1997.
- [33] CABRELLI, CARLOS, HEIL, CHRISTOPHER and MOLTER, URSULA. *Accuracy of Lattice Translates of Several Multidimensional Refinement Functions*. Jour. Approx. Theory, 95, 5–52. 1999.
- [34] CABRELLI, CARLOS, HEIL, CHRISTOPHER and MOLTER, URSULA. *Self-Similarity and Multiwavelets in Higher Dimensions*. Technical Report, School of Mathematics, Georgia Institute of Technology. 1999. Submitted.
- [35] CALDERÓN, A.P. *Intermediate Spaces and Interpolation, the Complex Method*. Stud. Math., 24, 113–190. 1964.
- [36] CALDERÓN, A.P. and ZYGMUND, A. *On Singular Integrals*. American Journ. Math., 78, 289–320. 1956.
- [37] CALDERÓN, A.P. and ZYGMUND, A. *Singular Integral Operators and Differential Equations*. American Journ. Math., 79, 901–921. 1957.
- [38] CAREY, W. K., HEMAMI, S. S. and HELLER, P. N. *Smoothness-Constrained Wavelet Image Compression*. Proceedings Int. Conf. Image Proc. 1996, 2, 6–10. Lausanne, Switzerland. 1996.
- [39] CATTÉ, F., LIONS, P.-L., MOREL, J.-M. and COLL, T. *Image Selective Smoothing and Edge Detection by Nonlinear Diffusion*. SIAM Journal on Numerical Analysis, 29, 182–193. 1992.
- [40] CARNICER, J.M., DAHMEN, W. and PEÑA, J.M.. *Local Decompositions of Refinable Spaces*. Appl. Comp. Harm. Analysis, 3, 127–153. 1996.
- [41] CHAMBOLLE, A., DEVORE, R. A., LEE, N.-Y. and LUCIER, B. *Nonlinear Wavelet Image Processing: Variational Problems, Compression and Noise Removal through Wavelet Shrinkage*. IEEE Trans. Image Proc., 7, 319–335. 1998.
- [42] CHUI, CHARLES K. *An Introduction to Wavelets*. Wavelet Analysis and Its Applications 1. ISBN 0-12-174584-8. Academic Press, Inc., Boston – San Diego – New York. 1992.
- [43] CHUI, CHARLES K., (editor). *Wavelets: A Tutorial in Theory and Applications*. Wavelet Analysis and Its Applications 2. ISBN 0-12-174590-2. Academic Press, Inc., Boston – San Diego – New York, 1992.

- [44] COHEN, ALBERT. *Ondelettes, analyses multirésolutions et filtres miroirs en quadrature*. Ann. Inst. H. Poincaré, Analyse Non Linéaire, 7, 439–459. 1990.
- [45] COHEN, ALBERT. *Biorthogonal Wavelets*. **In:** An Introduction to Wavelets [42], 123–152. ISBN 0-12-174584-8. 1992.
- [46] COHEN, A. and DAUBECHIES, I. *Orthonormal bases of compactly supported wavelets III. Better frequency resolution*. SIAM J. Math. Anal., 24(2), 520–527. 1993.
- [47] COHEN, ALBERT and DAUBECHIES, INGRID. *Non-separable Bidimensional Wavelet Bases*. Revista Matematica Iberoamericana, 9(1), 51–137. 1993.
- [48] COHEN, ALBERT and DAUBECHIES, INGRID. *A New Technique to Estimate the Regularity of Refinable Functions*. Revista Matematica Iberoamericana, 12, 527–591. 1996.
- [49] COHEN, A., DAUBECHIES, I. and FEAUVEAU, J. *Bi-orthogonal bases of compactly supported wavelets*. Comm. Pure and Appl. Math., 45, 485–560. 1992.
- [50] COHEN, A., DAUBECHIES, I. and VIAL, P. *Wavelets and Fast Wavelet Transform on the Interval*. Appl. Comp. Harm. Analysis, 1, 54–82. 1993.
- [51] COIFMAN, RONALD R., and WEISS, GUIDO. *Extensions of Hardy Spaces and Their Use in Analysis*. Bull. Amer. Math. Soc., 83, 569–645. 1977.
- [52] COIFMAN, RONALD R., MEYER, YVES and WICKERHAUSER, M. VICTOR. *Wavelet Analysis and Signal Processing*. **In:** Wavelets and Their Applications [199] 1992.
- [53] COTTET, G. H. and GERMAIN, L. *Image Processing Through Reaction Combined with Nonlinear Diffusion*. Math. Comp., 61, 659–673 1993.
- [54] COXETER, H.S.M.. *Regular Polytopes*. Dover, New York. 1973.
- [55] CROISIER, A., ESTEBAN, D. and GALAND, C. *Perfect Channel Splitting by Use of Interpolation/Decimation/Tree Decomposition Techniques*. **In:** Int. Conf. on Inform. Sciences and Systems, 443–446. Patras, Greece. 1976.
- [56] DAHMEN, WOLFGANG. *Wavelet and Multiscale Methods for Operator Equations*. Acta Numerica, 6, 55–228. 1997.
- [57] DAHMEN, W. and KUNOTH, A. *Multilevel Preconditioning*. Numerical Mathematics, 63, 315–344. 1992.
- [58] DAHMEN, W. and SCHNEIDER, R. *Wavelets on Manifolds I. Construction and Domain Decomposition*. SIAM J. Math. Anal., 31, 184–231. 1999.
- [59] DARWIN, CHARLES. *On the Origin of Species*. 1859.
- [60] DAUBECHIES, I. *Orthonormal Bases of Compactly Supported Wavelets*. Comm. Pure and Appl. Math., 41, 909–996. 1988.
- [61] DAUBECHIES, I. *The Wavelet Transform, Time-Frequency Localization and Signal Analysis*. IEEE Trans. on Inf. Theory, 36(5), 961–1005. 1990.
- [62] DAUBECHIES, INGRID. *Ten Lectures on Wavelets*. ISBN 0-89871-274-2. CBMS-NSF Regional Conference Series in Applied Mathematics No. 61. SIAM Publishing, Philadelphia. 1992.

- [63] DAUBECHIES, INGRID. *Orthonormal bases of compactly supported wavelets: II. Variations on a theme*. SIAM J. Math. Anal., 24, 499–519. 1993.
- [64] DAUBECHIES, INGRID. *Wavelet transforms and orthonormal wavelet bases*. **In:** Different Perspectives on Wavelets, DAUBECHIES, INGRID, (editor). Proceedings of Symposia in Applied Mathematics 47, 1–33. American Math. Soc., Providence, RI. 1993.
- [65] DAUBECHIES, I., GUSKOV, I., SCHRÖDER, P. and SWELDENS, W.. *Wavelets on Irregular Point Sets*. Phil. Trans. R. Soc. London A, 357(1760), 2397–2413. 1999.
- [66] DAUBECHIES, I., GUSKOV, I. and SWELDENS, W.. *Commutation for Irregular Subdivision*. Preprint, Bell Laboratories, Lucent Technologies. 1998.
- [67] DAUBECHIES, I., HAN, B., RON, A. and SHEN, Z. *Framelets: MRA-based Constructions of Wavelet Frames*. Preprint. 2001.
- [68] DAUBECHIES, INGRID and LAGARIAS, J. *Two-scale difference equations I. Global regularity of solutions*. SIAM J. Math. Anal., 22(5), 1388–1410. 1991.
- [69] DAUBECHIES, I. and LAGARIAS, J.. *Two-scale difference equations II. Local regularity, infinite products of matrices and fractals*. SIAM J. Math. Anal., 23(4), 1031–1079. 1992.
- [70] DAUBECHIES, INGRID and SWELDENS, WIM. *Factoring Wavelet Transforms into Lifting Steps*. Technical Report, Bell Laboratories, Lucent Technologies. 1996.
- [71] DE GIORGI, E. and FRANZONI, T. *Su un Tipo di Convergenza Variazionale*. Ren. Sem. Mat. Brescia, 3, 63–101. 1979.
- [72] DESLAURIERS, G. and DUBUC, S. *Interpolation Dyadique*. **In:** Fractals, Dimensions non Entières et Applications, CHERBIT, G. (editor), 44–55. Masson, Paris. 1987.
- [73] DEVORE, R., JAWERTH, B. and POPOV, V. *Compression of Wavelet Decompositions*. American Jour. Math., 114, 737–785. 1992.
- [74] DEVORE, R. and LUCIER, B. J. *Wavelets*. Acta Numerica, 1 1–56. 1991.
- [75] DEVORE, R. and LUCIER, B. J. *Classifying the Smoothness of Images: Theory and Applications to Wavelet Image Processing*. Proceedings Int. Conf. Image Proc. 1994, 2, 6–10. Los Alamitos, California. 1994.
- [76] DONOHO, DAVID L. *Interpolating Wavelet Transforms*. Preprint, Department of Statistics, Stanford University, 1992.
- [77] DONOHO, DAVID L. and JOHNSTONE, IAN. *Ideal Spatial Adaption by Wavelet Shrinkage*. Biometrika, 81, 425–455. 1994.
- [78] DONOHO, D., JOHNSTONE, I., KERKYACHARIAN, G. and PICARD, D. *Wavelet Shrinkage: Asymptopia?*. Journal of the Royal Statistical Society, Series B, 57, 301–369. 1995.
- [79] DUFFIN, R. J. and SCHAEFFER, A. C. *A Class of Nonharmonic Fourier Series*. Trans. Amer. Math. Soc., 72, 341–366. 1952.
- [80] EBERLY, DAVID. *A Differential Geometric Approach to Anisotropic Diffusion*. **In:** Geometry-Driven Diffusion in Computer Vision [102], 371–392. 1994.

- [81] EIROLA, T.. *Sobolev Characterization of Solutions of Dilation Equations*. SIAM Jour. Math. Analysis, 23(4), 1015–1030. 1992.
- [82] ERLEBACHER, GORDON and HUSSAINI, M. YOUSSEFF and JAMESON, LELAND M. *Wavelets — Theory and Applications*. ICASE/LaRC Series in Computational Science and Engineering. Oxford University Press, New York — Oxford. 1996.
- [83] FARID, H. and SIMONCELLI, E. P. *Optimally Rotation-Equivariant Directional Derivative Kernels*. In: Computer Analysis of Images and Patterns CAIP 1997, G. SOMMER, K. DANILIDIS and J. PAULI (eds.). Lecture Notes in Computer Science, 1296. Springer Verlag, Berlin, 207–214. 1997.
- [84] FELSBERG, MICHAEL and SOMMER, GERALD. *Scale Adaptive Filtering Derived from the Laplace Equation*. Jahrestagung der Deutschen Arbeitsgemeinschaft für Mustererkennung (DAGM) 2001, München. Springer Verlag, Berlin — Heidelberg — New York. 2001.
- [85] FIX, G. and STRANG, G.. *Fourier Analysis of the Finite Element Method in Ritz-Galerkin Theory*. Stud. Appl. Math., 48, 265–273. 1969.
- [86] FLEET, DAVID J. and LANGLEY, KEITH. *Recursive Filters for Optical Flow*. IEEE Trans. Pattern Analysis and Machine Intelligence, 17(1): 61–67. 1995.
- [87] FOURIER, JEAN BAPTISTE. *The Analytical Theory of Heat*. London, Cambridge University Press, 1878. First Published 1822.
- [88] FRANKLIN, PH.. *A Set of Continuous Orthogonal Functions*. Math Ann., 100, 522–529. 1928.
- [89] GABOR, D. *Theory of Communication*. J. Inst. Electr. Eng., London, 93(III): 429–457. 1946.
- [90] GALVIN, B., MCCANE, B., NOVINS, K., MASON, D. and MILLS, S. *Recovering Motion Fields: An Analysis of Eight Optical Flow Algorithms*. Proceedings 1998 British Machine Vision Conference (BMVC '98), 195–204. Southampton, United Kingdom. 1998.
- [91] GEMAN, D. and YANG, C. *Nonlinear Image Recovery with Half-Quadratic Regularization*. IEEE Trans. Pattern Analysis and Machine Intelligence, 14, 367–383. 1995.
- [92] GLOWINSKI, R., LAWTON, W., RAVACHOL, M. and TENENBAUM, E. *Wavelet Solution of Linear and Nonlinear Elliptic, Parabolic and Hyperbolic Problems in One Dimension*. In: GLOWINSKI, R. and LICHNEWSKI, A. (eds.) Proceedings of the 9th Int. Conf. on Computing Methods in Applied Sciences and Engineering. SIAM Press, Philadelphia. 1990.
- [93] GOUSSEAU, YANN and MOREL, JEAN-MICHEL. *Are Natural Images of Bounded Variation?* SIAM Journal on Mathematical Analysis, 33(3), 634–648. 2001.
- [94] GRANLUND, GÖSTA H. and KNUTSSON, HANS. *Signal Processing for Computer Vision*. Kluwer Academic Publishers. Dordrecht — Boston — London. ISBN 0-7923-9530-1. 1995.
- [95] GRIPENBERG, GUSTAF. *Computing the Joint Spectral Radius*. Linear Algebra and its Applications, 234, 43–60. 1996.
- [96] GRÖCHENIG, K. *Analyse Multiéchelle et Bases d'Ondelettes*. C.R. Acad. Sci. Paris, 305, Série I, 13–17. 1987.

- [97] GRÖCHENIG, K. and MADYCH, W.R. *Multiresolution Analysis, Haar Bases, and Self-Similar Tilings of \mathbb{R}^n* . IEEE Trans. on Inf. Theory, 38(2), 556–574. 1992.
- [98] GRÖCHENIG, KARLHEINZ and RON, AMOS. *Tight Compactly Supported Wavelet Frames of Arbitrarily High Smoothness*. Proc. American Mathematical Society, to appear. 1999.
- [99] GROSSMANN, A., MORLET, J. and PAUL, T. *Transforms Associated to Square Integrable Group Representations I, General Results*. Jour. Math. Phys., 27, 2473–2479. 1985.
- [100] GUSKOV, IGOR *Multivariate Subdivision Schemes and Divided Differences*. Preprint, Princeton University. 1998.
- [101] HAAR, ALFRED. *Zur Theorie der orthogonalen Funktionen-Systeme*. Math. Annalen, 69, 331–371. 1910.
- [102] TER HAAR ROMENY, BART M. (editor). *Geometry-Driven Diffusion in Computer Vision*. Computational Imaging and Vision 1. ISBN 0-7923-3087-0. Kluwer Academic Publishers, Dordrecht. 1994.
- [103] HACKBUSCH, WOLFGANG. *Iterative Solution of Large Sparse Systems of Equations*. Applied Mathematical Sciences. Springer-Verlag, Berlin — New York — Heidelberg. 1994.
- [104] HALMOS, PAUL R. *I Want to Be a Mathematician: An Automatology in Three Parts*. ISBN 0-88385-445-7. Springer-Verlag, Berlin — New York — Heidelberg. 1985.
- [105] HAN, BIN. *Analysis and Construction of Optimal Multivariate Biorthogonal Wavelets with Compact Support*. SIAM J. Math. Anal., 31, 274–304. 2000.
- [106] HAN, BIN and JIA, RONG-QING. *Quincunx Fundamental Refinable Functions and Quincunx Biorthogonal Wavelets*. Math. Comp. To appear. 2000.
- [107] HAN, BIN and JIA, RONG-QING. *Multivariate Refinement Equations and Subdivision Schemes*. SIAM J. Math. Anal. To appear. 2000.
- [108] HAUSSECKER, HORST W. and FLEET, DAVID J. *Computing Optical Flow with Physical Models of Brightness Variation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(6), 661–673. 2001.
- [109] HEIL, CHRISTOPHER and WALNUT, DAVID. *Continuous and Discrete Wavelet Transforms*. SIAM Review, 31, 628–666. 1989.
- [110] HEIL, CHRISTOPHER and COLELLA, DAVID. *Dilation Equations and the Smoothness of Compactly Supported Wavelets*. In: Wavelets: Mathematics and Applications [18]. 1994.
- [111] HERNÁNDEZ, EUGENIO and WEISS, GUIDO. *A First Course on Wavelets*. ISBN 0-8493-8274-2. Studies in Advanced Mathematics. CRC Press, Boca Raton — New York — London — Tokyo. 1996.
- [112] HOLSCHNEIDER, MATTHIAS and PINKALL, ULRICH. *Quadrature Mirror Filters and Loop Groups*. Technical Report CPT-94 P 3017. University of Marseille, France. 1994.
- [113] HOLSCHNEIDER, MATTHIAS. *Wavelet Analysis Over Abelian Groups*. Journal Applied Computational Harmonic Analysis, 2, 52–60. 1995.

- [114] HORN, B. and SCHUNCK, B. *Determining Optical Flow*. Artificial Intelligence, 17, 185–203. 1981.
- [115] JAFFARD, STÉPHANE. *Wavelet Expansions, Function Spaces and Multifractal Analysis*. In: Twentieth Century Harmonic Analysis — A Celebration [30], pp. 127–144. 2000.
- [116] JANSSEN, A.J.E.M. *Gabor Representation of Generalized Functions*. Journal Math. Appl., 80, 377–394. 1981.
- [117] JIA, RONG-QING. *Characterization of Smoothness of Multivariate Refinable Functions in Sobolev Spaces*. Trans. Amer. Math. Soc., 351, 4089–4112. 1999.
- [118] JOURNÉ, J. L. *Calderón-Zygmund Operators, Pseudo-Differential Operators and the Cauchy Integral of Calderón*. Lecture Notes in Mathematics, 994. Springer-Verlag, Berlin — New York — Heidelberg. 1983.
- [119] KALISA, C. and TORRÉSANI, B. *N-Dimensional Affine Weyl-Heisenberg Wavelets*. Ann. Inst. H. Poincaré, 59, 201–236. 1993
- [120] KICHENASSAMY, SATYANAD. *The Perona-Malik Paradox*. SIAM Journal Appl. Mathematics, 57(5), 1328–1342. 1997
- [121] KIMMEL, R. and SOCHEN, N. *Orientation Diffusion or How to Comb a Porcupine?* Technical Report CIS–2000–02. Computer Science Department, Technion, Haifa, Israel. 2000
- [122] KIMMEL, R., MALLADI, R. and SOCHEN, N. *Image Processing via the Beltrami Operator: Algebraic Frames for the Perception-Action Cycle* 2000. Lecture Notes in Computer Science, 1888, 232–239. Springer Verlag, Berlin — Heidelberg — New York. 2000
- [123] KIMMEL, R., SOCHEN, N. and MALLADI, R. *From High Energy Physics to Low Level Vision*. First Int. Conference on Scale-Space Theory in Computer Vision. Lecture Notes in Computer Science, 1252, 236–247. Springer Verlag, Berlin — Heidelberg — New York. 1997
- [124] KOENDERINK, J. J. *The Structure of Images*. Biological Cybernetics, 50, 363–370. 1984.
- [125] KÖRNER, THOMAS W. *Fourier Analysis*. Cambridge University Press, Cambridge. 1988.
- [126] KOVAČEVIĆ, J. and SWELDENS, W. *Wavelet Families of Increasing Order in Arbitrary Dimensions*. IEEE Trans. on Image Process. 1999. Submitted.
- [127] KOVAČEVIĆ, J. and VETTERLI, M. *Nonseparable Multidimensional Perfect Reconstruction Filter Banks and Wavelet Bases for \mathbb{R}^n* . IEEE Trans. Inform. Th., special issue on Wavelet Transforms and Multiresolution Signal Analysis, 38(2), 533–555. 1992.
- [128] KOVAČEVIĆ, J. and VETTERLI, M. *FCO Sampling of Digital Video Using Perfect Reconstruction Filter Banks*. IEEE Trans. on Image Process., 2(1), 118–122. 1993.
- [129] KOVAČEVIĆ, J. and VETTERLI, M. *Nonseparable Two- and Three-Dimensional Wavelets*. IEEE Trans. on Signal Process., 43(5), 1269–1273. 1995.

- [130] KUNOTH, ANGELA. *Multilevel Preconditioning*. Phd thesis, Freie Universität Berlin, Berlin, 1994.
- [131] KUNOTH, ANGELA and PIQUEMAL, ANNE-SOPHIE and VORLOEPER, JÜRGEN. *Multilevel Preconditioners for Discretizations of Elliptic Boundary Value Problems — Experiences with Different Software Packages*. IGPM Preprint 194. RWTH Aachen. 2000.
- [132] LATTO, A., RESNIKOFF, H.L. and TENENBAUM, E. *The Evaluation of Connection Coefficients of Compactly Supported Wavelets*. French – USA Workshop on Wavelets and Turbulence. Princeton, New Jersey. 1991.
- [133] LAU, KA-SING, (editor). *Advances in Wavelets*. ISBN 981-4021-08-3. Springer-Verlag, Singapore — Berlin — New York — Heidelberg. 1998.
- [134] LAWTON, W. M. *Necessary and Sufficient Conditions for Constructing Orthonormal Wavelet Bases*. Jour. Math. Phys, 32, pp. 57–61. 1991.
- [135] LAWTON, W. M. and RESNIKOFF, H. L. *Fractal Tiling for Multiple Mirror Optical Devices*. United States Patent 4,904,073. 1990.
- [136] LAWTON, W. M. and RESNIKOFF, H. L. *Multidimensional Wavelet Bases*. Aware Report AD910130, Aware, Inc., Cambridge, MA-02138. 1991.
- [137] LEMARIÉ, P. G. and MEYER, Y. *Ondelettes et Bases Hilbertiennes*. Rev. Mat. Iberoamericana, 2, 1–18. 1986.
- [138] LINDBERG, T. *Scale Space Theories in Computer Vision*. Kluwer Academic Publishers, Dordrecht — Boston — London. ISBN 0-7923-9418-6 1994.
- [139] LITTLEWOOD, J. E. and PALEY, R. E. A. C. *Theorems on Fourier Series I and II*. J. London Math. Soc., 6, 230–233. 1931. Proc. London Math. Soc., 42, 52–89. 1936.
- [140] LOUIS, ALFRED KARL, MAASS, PETER and RIEDER, ANDREAS. *Wavelets*. ISBN 3-519-02094-7. Teubner Studienbücher Mathematik. B.G. Teubner, Stuttgart. 1994.
- [141] LOW, F.. *Complete Sets of Wave Packets*. A Passion for Physics — Essays in Honor of Geoffrey Chew. World Scientific, 17–22. 1985.
- [142] MAASS, PETER. *Families of Orthogonal 2d Wavelets*. SIAM J. Math. Anal., 56. 1996.
- [143] MADYCH, W. R. *Some Elementary Properties of Multiresolution Analyses of $L^2(\mathbb{R}^n)$* . In: *Wavelets: A Tutorial in Theory and Applications* [43]. 1992.
- [144] MALLAT, STÉPHANE G. *Multiresolution Approximation and Wavelets*. Trans. Amer. Math. Soc., 315, 69–88. 1989.
- [145] MALLAT, STÉPHANE G. *A theory for Multiresolution Signal Decomposition: The Wavelet Representation*. IEEE Trans. Pattern Anal. Machine Intell., 11(7), 674–693. 1989.
- [146] MALLAT, STÉPHANE G. *Zero Crossings of a Wavelet Transform*. IEEE Trans. on Inf. Theory, 37(4), 1019–1033. 1991.
- [147] MALLAT, STÉPHANE G. *A Wavelet Tour of Signal Processing*. ISBN 0-12-466605-1. Academic Press, San Diego — London — Boston — New York — Sydney — Tokyo — Toronto. 1998.

- [148] MALLAT, STÉPHANE and HWANG, W. L. *Singularity detection and processing with wavelets*. IEEE Trans. Inform. Theory, 38, 617–643. 1992.
- [149] MALLAT, STÉPHANE and ZHONG, SIFEN. *Characterization of Signals from Multiscale Edges*. IEEE Trans. Pattern Analysis and Machine Intelligence, 14(7), 710–732. 1992.
- [150] MANN, THOMAS. *Der Zauberberg*. 1924.
- [151] MARQUARDT, D. W. *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*. Journal Soc. Industrial and Applied Math., 11, 431–441. 1963.
- [152] MARR, D. and HILDRETH, E. *Theory of Edge Detection*. Proceedings Royal Soc. London Series B, 207, 187–217. 1980.
- [153] MARR, DAVID. *Vision*. W.H. Freeman and Company, San Francisco. 1982.
- [154] MCCLELLAN, J. *The Design of Two-Dimensional Filters by Transformation*. Proc. Seventh Annual Princeton Conference on ISS, 247–251. 1973.
- [155] MERTINS, ALFRED. *Signaltheorie*. ISBN 3–519–06178–3. B.G. Teubner, Stuttgart. 1996.
- [156] MEYER, YVES, (editor). *Wavelets and Applications*. Number 20 in Research notes in Applied Mathematics. Springer Verlag, Berlin — Heidelberg — New York. 1991.
- [157] MEYER, YVES. *Wavelets and Operators*. ISBN 0–521–42000–8. Number 37 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge. 1992.
- [158] MEYER, YVES. *Wavelets: Algorithms and Applications*. SIAM, Philadelphia. 1993.
- [159] MEYER, YVES and COIFMAN, RONALD R. *Wavelets — Calderón-Zygmund and Multilinear Operators*. ISBN 0–521–42001–6. Number 48 in Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge. 1997.
- [160] MITREA, M. *Singular Integrals, Hardy Spaces and Clifford Wavelets*. Lecture Notes in Mathematics, 1575. Springer Verlag, Berlin – Heidelberg – New York. 1994.
- [161] MONZÓN, LUCAS, BEYLKIN, GREGORY and HEREMAN, WILLY. *Compactly Supported Wavelets Based on Almost Interpolating and Nearly Linear Phase Filters (Coiflets)*. Jour. Applied and Computational Harmonic Analysis, 7, 184–210. 1999.
- [162] MUMFORD, DAVID and SHAH, JAYANT. *Boundary Detection by Minimizing Functionals I*. Proceedings IEEE Conference on Computer Vision and Pattern Recognition CVPR '85, 22–26. San Francisco, United States of America. IEEE Computer Society Press, Washington. 1985.
- [163] MUMFORD, DAVID and SHAH, JAYANT. *Optimal Approximations by Piecewise Smooth Functions and Associated Variational Problems*. Communications on Pure and Applied Mathematics, 42, 577–685. 1989.
- [164] MUNCH, J. *Noise Reduction in Tight Weyl-Heisenberg Frames*. IEEE Trans. Inform. Th., special issue on Wavelet Transforms and Multiresolution Signal Analysis, 38(2), 608–616. 1992.

- [165] NAGEL, H.-H. and ENKELMANN, W. *An Investigation of Smoothness Constraints for the Estimation of Displacement Vector Fields from Image Sequences*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 8, 565–593. 1986.
- [166] NAKAYAMA, K. and SILVERMAN, G. H. *The Aperture Problem I: Perception of Nonrigidity and Motion Direction in Translating Sinusoidal Lines*. Vision Research, 28, 739–746. 1988.
- [167] NAKAYAMA, K. and SILVERMAN, G. H. *The Aperture Problem II: Spatial Integration of Velocity Information Along Contours*. Vision Research, 28, 747–753. 1988.
- [168] NECKELS, KAI. *An Alternative Approach to the Quaternionic Fourier Transform*. Internal Report. 1999.
- [169] NECKELS, KAI. *Wavelet Filter Design via Linear Independent Basic Filters*. In: Algebraic Frames for the Perception Action Cycle 2000, G. SOMMER and Y. Y. ZEEVI (eds.). Lecture Notes in Computer Science, 1888. Springer Verlag, Berlin, 251–258. 2000.
- [170] NECKELS, KAI. *Wavelet Connection Coefficients in Higher Dimensions*. Lecture held at Klausurtagung des Graduiertenkollegs "Effiziente Algorithmen und Mehrskalenmethoden". Malente. 2000.
- [171] NECKELS, KAI. *Tight Wavelet Frames of Arbitrary Smoothness in Higher Dimensions*. Paper Draft. 2001.
- [172] NECKELS, KAI. *Fast Local Estimation of Optical Flow Using Wavelet and Variational Methods*. In: Computer Analysis of Images and Patterns CAIP 2001, W. SKARBEEK (editor). Lecture Notes in Computer Science, 2124. Springer Verlag, Berlin, 349–356. 2001.
- [173] NEGAHDARIPOUR, SHAHRIAR. *Revised Definition of Optical Flow: Integration of Radiometric and Geometric Clues for Dynamic Scene Analysis*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(9), 961–979. 1998.
- [174] NEGAHDARIPOUR, S. and YU, C. H. *A Generalized Brightness Change Model for Optical Flow*. Proceeding Int. Conference on Computer Vision, 2–11. Berlin — Germany. 1993.
- [175] NEUMANN, JOHN VON. *Mathematical Foundations of Quantum Mechanics*. Princeton University Press, Princeton. 1932.
- [176] VON OEHSSEN, MARKUS. A proof for the consistency of AOS schemes. Private communication. 2001.
- [177] PALEY, R. E. A. C. *A Remarkable System of Orthogonal Functions*. Proc. London Math. Soc., 34, 241–279. 1932.
- [178] PALEY, R. E. A. C. and ZYGMUND, A. *On Some Series of Functions*. Proc. Cambridge Phil. Soc., 28, 190–205. 1932.
- [179] PAZY, AMNON. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Applied Mathematical Sciences 44. Springer Verlag, New York — Heidelberg — Berlin. 1983.
- [180] PERONA, P. and MALIK, J. *Scale Space and Edge Detection Using Anisotropic Diffusion*. IEEE Trans. Pattern Analysis and Machine Intelligence, 12, 629–639. 1990.

- [181] PERONA, P., SHIOTA, T. and MALIK, J. *Anisotropic Diffusion*. **In:** Geometry-Driven Diffusion in Computer Vision [102], 73–92. 1994.
- [182] PROESMANS, M., PAUWELS, E. J. and GOOL, L.J. VAN. *Coupled Geometry-Driven Diffusion Equations for Low-Level Vision*. **In:** Geometry-Driven Diffusion in Computer Vision [102], 191–228. 1994.
- [183] RENARDY, MICHAEL and ROGERS, ROBERT C. *An Introduction to Partial Differential Equations*. Texts in Applied Mathematics 13. ISBN 0–387–97952–2. Springer-Verlag, Berlin – Heidelberg – New York. 1993.
- [184] RESNIKOFF, HOWARD L. and WELLS, RAYMOND O. *Wavelet Analysis — The Scalable Structure of Information*. ISBN 0–387–98383–X. Springer-Verlag, Berlin – Heidelberg – New York. 1998.
- [185] RICHARDSON, THOMAS L. *Limit Theorems for a Variational Problem Arising in Computer Vision*. Annali della Scuola Normale, 19, Fasc. 1. 1992.
- [186] RICHTER, GOTTHARD M. *Digitale Bildverarbeitung in der Astronomie*. Autbild '81. 1. Jenaer Symposium zur automatischen Bildverarbeitung. Wissenschaftliche Beiträge der Friedrich-Schiller-Universität Jena.. 1981.
- [187] RICHTER, GOTTHARD M. *Verfahren tur Bildtransformation*. Autbild '84/1. INQUAMESS-Lehrgang Automatische Bildverarbeitung, Kurs 1, Wandlitzsee. Wissenschaftliche Beiträge der Friedrich-Schiller-Universität Jena.. 1984
- [188] RIEDER, A., WELLS, R.O., JR. and ZHOU, X. *A Wavelet Approach to Robust Multilevel Solvers for Anisotropic Elliptic Problems*. Applied and Comp. Harmonic Analysis, 1, 355–367. 1994.
- [189] RIEMENSCHNEIDER, S.D. and SHEN, Z. *Wavelets and Prewavelets in Low Dimensions*. Jour. Approx. Theory, 71, 18–38. 1992.
- [190] RIESZ, R. and SZ-NAGY, B. *Functional Analysis*. Frederick Ungar Publishing, New York. 1955.
- [191] RIOUL, O. *Simple Regularity Criteria for Subdivision Schemes*. SIAM Jour. Math. Analysis, 23, 1544–1576 1992.
- [192] RON, AMOS and SHEN, ZUOWEI. *Weyl-Heisenberg Frames and Riesz Bases in $L^2(\mathbb{R}^d)$* . Duke Mathematical Journal, 89, 237–282. 1997.
- [193] RON, AMOS and SHEN, ZUOWEI. *Affine Systems in $L^2(\mathbb{R}^d)$: The Analysis of the Analysis Operator*. Jour. Functional Analysis, 148, 408–447. 1997.
- [194] RON, AMOS and SHEN, ZUOWEI. *Compactly Supported Tight Affine Spline Frames in $L^2(\mathbb{R}^d)$* . Math. Comp., 67, 191–207. 1998.
- [195] RON, AMOS and SHEN, ZUOWEI. *Construction of Compactly Supported Affine Frames in $L^2(\mathbb{R}^d)$* . **In:** Advances in Wavelets [133], 27–49. 1998.
- [196] RON, AMOS and SHEN, ZUOWEI. *The Sobolev Regularity of Refinable Functions*. Jour. Approx. Theory. To appear. 1999.

- [197] RUDIN, LEONID I. and OSHER, STANLEY. *Total Variation Based Image Restoration with Free Local Constraints*. Proc. IEEE Int. Conf. Image Processing ICIP-94, vol. 1, 31–35. IEEE Computer Society Press, Los Alamitos. 1994.
- [198] RUDIN, W. *Functional Analysis*. McGraw-Hill, New York. 1979.
- [199] RUSKAI, M.B., BEYLKIN, G., COIFMAN, R.R., DAUBECHIES, I., MALLAT, S., MEYER, Y. and RAPHAEL, L. (eds.) *Wavelets and Their Applications*. Jones and Bartlett, Boston. 1992.
- [200] SCHARR, H., KÖRKEL, S. and JÄHNE, B. *Numerische Isotropieoptimierung von FIR-Filtern Mittels Querglättung*. In: *Mustererkennung 1997*. WAHL, F. and PAULUS, E. (eds.), 367–374. Springer Verlag, Berlin. 2000.
- [201] SCHNÖRR, CHRISTOPH. *Unique Reconstruction of Piecewise-Smooth Images by Minimizing Strictly Convex Non-Quadratic Functionals*. Journal of Mathematical Imaging and Vision, 4, 189–198. 1994.
- [202] SCHNÖRR, CHRISTOPH and WEICKERT, JOACHIM. *Variational Image Motion Computation: Theoretical Framework, Problems and Perspectives*. In: *Mustererkennung 2000*. G. SOMMER, N. KRÜGER and C. PERWASS (eds.), 476–487. Springer Verlag, Berlin. 2000.
- [203] SEGMAN, J. and SCHEMPP, W. *Two Ways to Incorporate Scale in the Heisenberg Group With An Intertwining Operator*. Jour. Math. Imag. Vision, 3, 79–94. 1993.
- [204] SELESNICK, I.W., ODEGARD, J.E. and BURRUS, C.S. *Nearly Symmetric Orthogonal Wavelets with Non-Integer DC Group Delay*. In: *Proceedings of the Seventh Digital Signal Processing Workshop*, 413–434. Loen, Norway. 1996.
- [205] SHAH, JAYANT. *A Common Framework for Curve Evolution, Segmentation and Anisotropic Diffusion*. In: *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, 136–142. IEEE Computer Society Press, Los Alamitos. 1996.
- [206] SHAKESPEARE, WILLIAM. *King Lear*. 1607.
- [207] SIEGEL, CARL LUDWIG. *Vorlesungen über die Himmelsmechanik*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen 85. Springer Verlag, Heidelberg — Berlin. 1956.
- [208] SIEGEL, CARL LUDWIG. *Über einige Anwendungen Diophantischer Approximationen*. Gesammelte Abhandlungen I, 209–266. Springer Verlag, Heidelberg — Berlin. 1966.
- [209] SINGH, A. *An Estimation-Theoretic Framework for Image-Flow Computation*. Proc. IEEE Int. Conference on Computer Vision, 168–177. Osaka — Japan. 1990.
- [210] SMITH, M. J. T. and BARNWELL, T. P. *Exact Reconstruction for Tree-Structured Subband Coders*. IEEE Trans. on Acoust. Signal Speech Process., 34(3), 431–441. 1986.
- [211] SPORRING, J. *The Entropy of Scale Space*. Proceedings 13th Int. Conference on Pattern Recognition (ICPR). Vienna, Austria. Vol. A, 900–904. 1996.
- [212] STANHILL, DAVID and ZEEVI, YEHOShUA Y. *Two-dimensional Orthogonal Wavelets with Vanishing Moments*. Technical Report EE 976. Department of Electrical Engineering at Technion, Haifa, Israel. 1995.

- [213] STRANG, GILBERT and NGUYEN, TRUONG. *Wavelets and Filter Banks*. ISBN 0-9614088-7-1. Wellesley-Cambridge Press, Wellesley, Massachusetts, 1997.
- [214] STRICHARTZ, RICHARD. *Construction of orthonormal wavelets*. **In:** *Wavelets: Mathematics and Applications* [18]. 1994.
- [215] STRÖMBERG, J.-O.. *A Modified Franklin System and Higher Order Spline Systems on \mathbb{R}^n as Unconditional Bases for Hardy Spaces*. **In:** *Conference on Harmonic Analysis in Honor of Antoni Zygmund II*, 475–494. BECKNER et. al. (eds.). Belmont, CA. 1983.
- [216] SWELDENS, W. *The lifting scheme: A custom-design construction of biorthogonal wavelets*. *Appl. Comput. Harmon. Anal.*, 3(2), 186–200. 1996.
- [217] SWELDENS, WIM and PIESSENS, ROBERT. *Quadrature Formulae and Asymptotic Error Expansions for Wavelet Approximations of Smooth Functions*. *SIAM Jour. Numerical Analysis*, 31(4), 1240–1264. 1994.
- [218] SWELDENS, WIM and SCHRÖDER, PETER. *Building Your Own Wavelets at Home*. **In:** *Wavelets in Computer Graphics*, ACM SIGGRAPH Course Notes. 1996.
- [219] TCHAMITCHIAN, PHILIPPE. *Wavelets, Functions and Operators*. **In:** *Wavelets — Theory and Applications* [82], 83–181. 1996.
- [220] TIKHONOV, A. N.. *Solution of Ill-Posed Problems*. V. H. Winston & Sons, John Wiley. 1977.
- [221] TRIEBEL, HANS *Theory of Function Spaces I*. Monographs in Mathematics, 78. Birkhäuser, Basel. 1983.
- [222] TRIEBEL, HANS *Theory of Function Spaces II*. Monographs in Mathematics, 84. Birkhäuser, Basel. 1984.
- [223] VAIDYANATHAN, P. P. *Multirate Systems and Filter Banks*. Prentice-Hall PTR, Englewood-Cliffs, New Jersey. 1993.
- [224] VAIDYANATHAN, P. P. and HOANG, P.-Q. *Lattice Structures for Optimal Design and Robust Implementation of Two-Channel Perfect Reconstruction Filter Banks*. *IEEE Trans. on Acoust. Signal Speech Process.*, 36(1), 81–94. 1988.
- [225] VETTERLI, MARTIN und HERLEY, CORMACK. *Wavelets and Filter Banks: Theory and Design*. *IEEE Trans. Signal Process.*, 40, 2207–2232. 1992.
- [226] VETTERLI, MARTIN und KOVAČEVIĆ, JELENA. *Wavelets and Subband Coding*. ISBN 0-13-097080-8. Prentice-Hall PTR, Englewood-Cliffs, New Jersey. 1995.
- [227] VILLEMOES, LARS F. *Energy Moments in Time and Frequency for Two-Scale Difference Equation Solutions and Wavelets*. *SIAM Jour. Math. Analysis* 23(6), 1057–1071. 1992.
- [228] VILLEMOES, LARS F. *Continuity of Nonseparable Quincunx Wavelets*. *Applied and Computational Harmonic Analysis* 1(2), 180–187. 1994.
- [229] WALSH, J. L. *A Closed Set of Normal Orthogonal Functions*. *Amer. Journ. Math.*, 43, 5–24. 1923.

- [230] WANG, YU-PING and LEE, S. L. *Scale-Space Derived From B-Splines*. IEEE Trans. on Pattern Recognition and Machine Intelligence, 20(10), 1040–1055. 1998.
- [231] WATSON, ANDREW B. und AHUMADA, ALBERT J. *A Hexagonal Orthogonal-Oriented Pyramid as a Model of Image Representation in Visual Cortex*. IEEE Transactions on Biomedical Engineering, 36(1), 97–106. 1989.
- [232] WATSON, ANDREW B. *Recursive In-Place Algorithm for the Hexagonal Orthogonal Oriented Quadrature Image Pyramid*. Advances in Image Compression and Automatic Target Recognition, 1099, 194–200. 1989.
- [233] WEBER, JOSEPH and MALIK, JITENDRA. *Robust Computation of Optical Flow in a Multi-Scale Differential Framework*. International Journal of Computer Vision, 2, 5–19. 1994.
- [234] WEICKERT, JOACHIM. *Foundations and Applications of Nonlinear Anisotropic Diffusion Filtering*. Z. Angew. Math. Mech., 76(1), 283–286. 1996.
- [235] WEICKERT, JOACHIM. *Nonlinear Diffusion Scale-Spaces: From the Continuous to the Discrete Setting*. In: ICAOS '96: Images, Wavelets and PDEs. BERGER, M. O., DERICHE, R., HERLIN, I., JAFFRÉ, J. and MOREL, J. M. (eds.). Lecture Notes in Control and Information Sciences, 219, 111–118. Springer Verlag, London — Berlin — Heidelberg. 1996.
- [236] WEICKERT, JOACHIM. *Anisotropic Diffusion in Image Processing*. European Consortium for Mathematics in Industry. ISBN 3–519–02606–6. B. G. Teubner, Stuttgart. 1998.
- [237] WEICKERT, JOACHIM and HAAR ROMENY, BART M. TER and VIERGEVER, MAX A. *Efficient and Reliable Schemes for Nonlinear Diffusion Filtering*. IEEE Transactions on Image Processing, 7(3): 398–410. 1998.
- [238] WEICKERT, JOACHIM and SCHNÖRR, CHRISTOPH. *Räumlich-zeitliche Berechnung des optischen Flusses mit nichtlinearen flußabhängigen Glattheitstermen*. In: Mustererkennung 1999. W. FÖRSTNER, J. M. BUHMANN, A. FABER and P. FABER (eds.), 317–324. Springer Verlag, Berlin. 1999.
- [239] WEISS, GUIDO and WILSON, E.N. *The Mathematical Theory of Wavelets*. In: Twentieth Century Harmonic Analysis — A Celebration [30], pp. 329–366. 2000.
- [240] WELLS, R. O. and ZHOU, X. *Wavelet interpolation and approximate solutions of partial differential equations*. In: Noncompact Lie Groups, WILSON, R. and TANNER, E. A., (eds.), No. 429 of Proceedings of NATO Advanced Research Workshop, 349–366. Kluwer Academic Publishers, Dordrecht – Boston – London. 1994.
- [241] WEYL, HERMANN. *Symmetry*. ISBN 0–691–08045–3. Princeton University Press. Princeton, New Jersey. 1952.
- [242] WHITAKER, R. T. and PIZER, S. M. *A Multi-Scale Approach to Nonuniform Diffusion*. Computer Vision, Graphics and Image Processing. Image Understanding, 57(1), 111–120. 1993.
- [243] WILSON, H. R. *Psychophysical Evidence for Spatial Channels*. In: Physical and Biological Processing of Images. BRADDICK, O. J. and SLEIGH, A. C., (eds). Springer Verlag, New York. 1983.

- [244] WITKIN, A. P. *Scale-Space Filtering*. Proceedings 8th International Joint Conference on Artificial Intelligence (IJCAI '83), 2, 1019–1022. Karlsruhe, Germany. 1983.
- [245] WOODS, JOHN W. *Subband Image Coding*. ISBN 0-7923-9093-8. Kluwer Academic Publishers, Boston — Dordrecht — London. 1991.
- [246] WOJTASZCZYK, PAVEL A *Mathematical Introduction to Wavelets*. London Mathematical Society Student Texts 37. ISBN 0-521-57894-9. Cambridge University Press, Cambridge. 1997.
- [247] WU, Y.-T., KANADE, T. COHN, J. and LI, C.-C. *Optical Flow Estimation Using Wavelet Motion Model*. Proceedings of the International Conference on Computer Vision, 992–998. 1998.
- [248] YOSIDA, K. *Functional Analysis*. Springer, New York — Heidelberg — Berlin. 1966.
- [249] YOUNG, R. A. *The Gaussian Derivative Model for Spatial Vision I. Retinal Mechanisms*. Spatial Vision, 2, 273–293. 1987.
- [250] <http://www.cs.otago.ac.nz/research/vision/>
- [251] <ftp://ftp.csd.uwo.ca/pub/vision/>

I did it my way.

—FRANK SINATRA